AIMS Mathematics, 11(8): 27164–27202.
DOI: 10.3934/math.20261088
Received: 30 June 2026
Revised: 16 August 2026
Accepted: 19 August 2026
Published: 28 August 2026

http://www.aimspress.com/journal/Math

Research article

Adaptive Koopman predictive control of large-scale interconnected nonlinear systems: A recursive lifting approach with small-gain guarantees

Tawfik Guesmi[1], **Mourad Elloumi**[2,3,*], **Khalid Alqunun**[1] **and Omar Naifar**[4,5]

[1] Department of Electrical Engineering, College of Engineering, University of Ha'il, Ha'il 55473, Saudi Arabia

[2] Faculty of Sciences of Gafsa, University of Gafsa, Gafsa, Tunisia

[3] Laboratory of Sciences and Technology of Automatic Control and Computer Engineering, National School of Engineering of Sfax, Sfax University, P.O. Box 1173, Sfax 3038, Tunisia

[4] Higher Institute of Applied Science and Technology of Kairouan, University of Kairouan, Kairouan, Tunisia

[5] Control and Energy Management Laboratory, National School of Engineering, Sfax University, Sfax, Tunisia

* **Correspondence:** Email: mourad.elloumi@enis.tn.

Abstract: Large-scale processes in the energy and process industries are networks of interconnected nonlinear subsystems with uncertain and slowly drifting dynamics. Block-oriented self-tuning schemes, such as Hammerstein-based regulators, handle this uncertainty recursively but tie the controller to a restrictive model structure. In this study, the identified block-oriented model is replaced by Koopman linear predictors estimated online, one per subsystem, in a lifted observable space through a recursive least-squares law with forgetting and bounded covariance. These adaptive predictors are integrated into a distributed, input-constrained Koopman model predictive controller, denoted AK-DMPC, whose local problems are convex quadratic programs solved in parallel with a single neighbor communication per sampling instant. Three results are established: (i) The identifier has bounded covariance, and its parameter error does not vanish but decays exponentially to a residual set whose radius is set by the lifting (closure) error and the parameter drift; (ii) the multistep prediction error is bounded by the estimation and closure errors propagated through the interconnection; and (iii) under a verifiable small-gain condition on the interconnection, the closed loop is input-to-state stable on an explicit region of the state space with respect to these errors. The network tracking error is therefore governed by the dictionary richness. Abrupt parameter jumps, which violate the slow-drift premise, are covered by a dwell-time corollary quantifying the excursion and the geometric recovery. Recursive feasibility is proved under standard terminal ingredients. On three interconnected nonlinear tanks with an abrupt valve fault, AK-DMPC lowers the network tracking error by 28% and 58% against a frozen Koopman controller

and a decentralized linear controller, and it matches an adaptive Hammerstein self-tuning regulator in accuracy with threefold smoother control and a closed-loop stability certificate ($\rho(A_{\mathrm{cl}}) \approx 0.54$).

Keywords: Koopman operator; adaptive control; distributed model predictive control; large-scale interconnected systems; data-driven control; recursive identification; input-to-state stability; small-gain theorem
Mathematics Subject Classification: 93C40, 93D25, 93B30, 93A15, 93C55

1. Introduction

The control of large-scale systems, such as power grids, water networks, chemical trains, and multi-stage separation units, is difficult because of their nonlinear dynamics; their many interconnected subsystems; and uncertain parameters that vary slowly because of fouling, aging, and load changes. Classical fixed-gain laws cannot keep good performance across such variations, which has motivated the development of adaptive and self-tuning control [1–3]. For the interconnected nonlinear case, block-oriented self-tuning regulators based on Hammerstein models have been developed and applied successfully: Recursive estimation of the static nonlinearity and of the linear dynamics are combined with a minimum-variance or pole-placement design, and convergence is established through prediction-error arguments [4–6]. However, this class of strategies suffers from a structural limitation, as the controller inherits the modeling bias of the chosen block structure; indeed, the static nonlinearity and linear dynamics factorization are rarely exact for strongly coupled processes.

The Koopman operator offers an attractive way to remove this structural restriction. It represents a nonlinear system through a linear, infinite-dimensional operator acting on observables of the state [7, 8]. Finite-dimensional approximations are then obtained by truncating the Koopman operator to a finite dictionary, which yields a finite linear predictor that can be identified from data by extended dynamic mode decomposition (EDMD) [9–11]; under sufficient enrichment of the dictionary, these approximations converge to the underlying operator [12, 13]. The surveys [14, 15] and the monograph [16] document the rapid maturation of the field, including finite linear representations tailored to control [17, 18] and identification through lifting [19, 20] or deep autoencoders [21–23]. As the truncated predictor is linear in the lifted coordinates, it meshes naturally with model predictive control (MPC): Korda and Mezić [24] introduced Koopman MPC, in which a convex quadratic program is solved over the lifted linear model, and this formulation has been followed by numerous variants [25–27], including neural network and multimodel extensions [28, 29], together with error analyses for prediction and closed-loop control [30]. The reach of the lifting idea now extends well beyond process systems. In aerospace applications, Chen and Shan [31] constructed a set of observables that renders the attitude dynamics on $SO(3)$ linear in the lifted coordinates and synthesized the attitude controller by linear optimal control on the resulting finite-dimensional model, thereby replacing a genuinely nonlinear synthesis with a linear one. More recently, Wang et al. [32] embedded a Koopman linear predictor inside a model predictive controller for the test mass capture maneuver of a drag-free satellite, where the lifted linear model is precisely what makes the constrained receding-horizon problem a tractable convex program. Both studies confirm, on demanding applications, the two premises on which the present work rests: That a modest state-inclusive dictionary can carry a strongly nonlinear plant and

that the lifted linear predictor is the natural interface to constrained predictive control. Both, however, identify the predictor *once, offline*, for a *single* rigid body or spacecraft, and neither addresses drift of the lifted model nor an interconnection of many such subsystems; they therefore delimit, rather than close, the two gaps identified below. In parallel, direct data-driven predictive control techniques based on Willems' fundamental lemma [33–35] have been developed to bypass the explicit model altogether.

Two gaps remain at the intersection of these developments, and they are precisely the ones relevant to the control of large-scale processes. First, the overwhelming majority of Koopman MPC formulations identify the predictor *offline* and keep it *frozen*. As a result, they inherit the very weakness that adaptive control was created to overcome, namely sensitivity to drift and faults. Although online and adaptive Koopman schemes are only now emerging [36–38], with wind farm and streaming data applications [39, 40], they are developed for single, centralized systems and come without closed-loop guarantees for interconnected plants. Second, the few Koopman predictive schemes that do provide stability and recursive feasibility guarantees [41] are likewise monolithic: They do not exploit or certify the interconnection structure that dominates large-scale systems for which distributed MPC is the established paradigm [42–44].

This paper closes both gaps by developing an *adaptive Koopman distributed predictive controller* (AK-DMPC) for large-scale interconnected nonlinear systems and by certifying it. The three contributions are the following.

1. We attach to each subsystem a Koopman linear predictor with explicit *cross-coupling* terms and estimate it *online* by a recursive least-squares law with forgetting and bounded covariance, the lifted-space counterpart of the constant-trace recursive estimators used in self-tuning control [2, 45]. We prove (Theorem 3.3) that the covariance is bounded and that the parameter error decays exponentially to a residual set—and, in general, to no smaller a set—whose radius is set by the closure error and the drift.

2. We bound the multistep prediction error of the resulting adaptive predictor over the network (Theorem 4.3), separating the estimation error, the closure error, and the interconnection coupling.

3. We prove that under a verifiable small-gain condition on the interconnection [46–48], the closed loop is input-to-state stable (ISS) [49–51] on an explicitly characterized region with respect to the closure/estimation errors (Theorem 4.8) so that the asymptotic network tracking error is governed by the dictionary richness; we also establish robust recursive feasibility (Proposition 4.11).

Around this core, we develop the operating envelope of the scheme: A dwell-time corollary that covers abrupt parameter jumps outside the slow-drift premise (Corollary 3.5), a quantitative account of how the identifier behaves when it is updated intermittently and when the data are noisy (Corollaries 3.7 and 3.8), a recursive update that penalizes the multistep—rather than only the one-step—prediction error (Section 4.1), and a systematic discussion of which standing assumptions can be relaxed in practice and at what price (Section 4.2). The scheme thus generalizes the block-oriented self-tuning regulators of [5, 6] (Hammerstein → Koopman; minimum variance → predictive; monolithic → distributed) while preserving their recursive, certainty-equivalence spirit. A benchmark of three interconnected nonlinear tanks [52, 53] with an abrupt valve fault validates the analysis.

The remainder of the paper is organized as follows. Section 2 fixes the notation, recalls the Koopman framework, and states the problem. Section 3 develops the adaptive Koopman identifier and Theorem 3.3,

together with its extensions to parameter jumps, intermittent updates, measurement noise and multistep-aware estimation. Section 4 develops the distributed predictive controller and the closed-loop guarantees (Theorems 4.3–4.8, Proposition 4.11) and discusses the standing assumptions. Section 5 reports the numerical study, and Section 6 concludes. Auxiliary proofs are collected in Appendices A–D.

2. Preliminaries and problem formulation

We write $\mathbb{R}$, $\mathbb{R}^n$, and $\mathbb{N}$ for the reals, the n-vectors, and the non-negative integers; $\|\cdot\|$ is the Euclidean norm for vectors and the induced spectral norm for matrices and $\|\cdot\|_F$ the Frobenius norm. For a symmetric matrix M, $M \succ 0$ ($M \succeq 0$) denotes positive (semi)definiteness, $\lambda_{\min}(M)$ and $\lambda_{\max}(M)$ its extreme eigenvalues, $\rho(M)$ its spectral radius, and $\mathrm{tr}(M)$ its trace; I_n is the identity. A continuous function $\alpha : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is of class $\mathcal{K}$ if it is strictly increasing with $\alpha(0) = 0$, and it is of class $\mathcal{K}_\infty$ if in addition it is unbounded; $\beta : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is of class $\mathcal{KL}$ if $\beta(\cdot, t) \in \mathcal{K}$ for each t and $\beta(s, t) \downarrow 0$ as $t \to \infty$ for each s [50, 51]. For a matrix $\Gamma \in \mathbb{R}^{N \times N}$ with non-negative entries, we say it is *Schur* if $\rho(\Gamma) < 1$; we use the standard fact that this is equivalent to the existence of a vector $v \succ 0$ with $\Gamma v \prec v$ [47].

We consider a large-scale system composed of N subsystems whose interconnection is described by a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with node set $\mathcal{V} = \{1, \ldots, N\}$ and neighbor sets $\mathcal{N}_i = \{\, j : (j, i) \in \mathcal{E} \,\}$. The ith subsystem evolves in discrete time as

$$x_i(k+1) = F_i\Big(x_i(k), u_i(k), \{x_j(k)\}_{j \in \mathcal{N}_i}\Big), \qquad i \in \mathcal{V}, \tag{2.1}$$

where $x_i \in \mathbb{X}_i \subset \mathbb{R}^{n_i}$ is the state, $u_i \in \mathbb{U}_i \subset \mathbb{R}^{m_i}$ the input, and $F_i : \mathbb{R}^{n_i} \times \mathbb{R}^{m_i} \times \mathbb{R}^{\sum_{j \in \mathcal{N}_i} n_j} \to \mathbb{R}^{n_i}$ an *unknown* continuous map. The aggregate state and input are $x = \mathrm{col}(x_1, \ldots, x_N)$ and $u = \mathrm{col}(u_1, \ldots, u_N)$. The sets $\mathbb{X}_i$ are compact, and $\mathbb{U}_i$ are compact and convex (e.g., box constraints). The objective is to drive each output x_i to a constant reference x_i^{r}, using only measured data and a single round of neighbor communication per step, with provable closed-loop guarantees.

The Koopman operator of an autonomous map $x^+ = F(x)$ acts on scalar observables ϕ by composition, $(\mathcal{U}\phi)(x) = \phi(F(x))$; it is linear but infinite-dimensional [7, 8, 14]. For a *controlled* and *interconnected* subsystem, the argument of the observable has to be enlarged accordingly. Writing $\chi_i \triangleq (x_i, u_i, \{x_j\}_{j \in \mathcal{N}_i}) \in \Omega_i$ for the extended variable of subsystem i with $\Omega_i \triangleq \mathbb{X}_i \times \mathbb{U}_i \times \prod_{j \in \mathcal{N}_i} \mathbb{X}_j$ its compact operating set, the *controlled Koopman operator* $\mathcal{U}_i$ of subsystem i acts on functions of χ_i by

$$(\mathcal{U}_i \phi)(\chi_i) \triangleq \phi\Big(F_i(x_i, u_i, \{x_j\}_{j \in \mathcal{N}_i})\Big); \tag{2.2}$$

that is, $\mathcal{U}_i$ maps a function of the current extended variable to the value of that function at the successor state. This is the standard extended-state (or "lifted-with-input") construction of Koopman control [17, 24, 54]; it is the operator that is truncated below. A finite dictionary of observables turns (2.2) into a finite linear predictor whose data-driven approximation is EDMD and whose error vanishes as the dictionary spans an invariant subspace [9, 12, 17]. We use this fact subsystemwise.

Definition 2.1. For each $i \in \mathcal{V}$, let $\Psi_i = \mathrm{col}(\psi_{i,1}, \ldots, \psi_{i,p_i}) : \mathbb{R}^{n_i} \to \mathbb{R}^{p_i}$ be a dictionary that is (i) *state-inclusive*: There exists a matrix $C_i \in \mathbb{R}^{n_i \times p_i}$ with $x_i = C_i \Psi_i(x_i)$ for all $x_i \in \mathbb{X}_i$; and (ii) *locally Lipschitz* on the compact set $\mathbb{X}_i$. That is, there exists $L_{\psi,i} > 0$ with $\|\Psi_i(x) - \Psi_i(y)\| \leq L_{\psi,i} \|x - y\|$ for all $x, y \in \mathbb{X}_i$. The lifted state is $z_i(k) \triangleq \Psi_i(x_i(k))$.

Lipschitz continuity (rather than mere continuity) is required for the explicit construction of the interconnection gains in Theorem 4.8; it holds for the polynomial, radial-basis, and smooth observables commonly used in EDMD [9, 55] and, for the dictionary of Section 5, $L_{\psi,i}$ is finite and computed in closed form.

State inclusion (achieved, e.g., by placing the components of x_i among the observables) guarantees exact output recovery and is standard in Koopman control [17, 24]. Because $\mathbb{X}_i$ is compact, and Ψ_i continuous, z_i is bounded; we write $\|z_i\| \le \bar{z}_i$ and $\|u_i\| \le \bar{u}_i$ on the operating set, and $\bar{z} \triangleq \max_i \bar{z}_i$, $\bar{u} \triangleq \max_i \bar{u}_i$.

Assumption 1. *There exist matrices $A_i^\star \in \mathbb{R}^{p_i \times p_i}$, $B_i^\star \in \mathbb{R}^{p_i \times m_i}$, $W_{ij}^\star \in \mathbb{R}^{p_i \times p_j}$ $(j \in \mathcal{N}_i)$ and a vector $e_i^\star \in \mathbb{R}^{p_i}$ such that along every admissible trajectory of* (2.1) *in $\prod_i \mathbb{X}_i \times \mathbb{U}_i$,*

$$z_i(k+1) = A_i^\star z_i(k) + B_i^\star u_i(k) + \sum_{j \in \mathcal{N}_i} W_{ij}^\star z_j(k) + e_i^\star + r_i(k), \qquad \sup_k \|r_i(k)\| \le \bar{r}_i, \tag{2.3}$$

where r_i is the closure (lifting) residual.

Assumption 1 is the interconnected counterpart of the standard premise of Koopman predictive control [24]. It is not vacuous: $\bar{r}_i$ is the *finite-dictionary projection error* of the controlled Koopman operator (2.2) onto the ansatz space actually used by the predictor. Because the right-hand side of (2.3) is affine in the input and in the neighbor lifts, that ansatz space is *not* the span of Ψ_i alone: It is the span of every regressor that (2.3) is allowed to use,

$$\mathcal{D}_i \triangleq \operatorname{span}\{\psi_{i,1}, \dots, \psi_{i,p_i}\} \oplus \operatorname{span}\{u_{i,1}, \dots, u_{i,m_i}\} \oplus \bigoplus_{j \in \mathcal{N}_i} \operatorname{span}\{\psi_{j,1}, \dots, \psi_{j,p_j}\} \oplus \operatorname{span}\{\mathbf{1}\}, \tag{2.4}$$

a subspace of $L^2(\Omega_i, \nu_i)$, where $\mathbf{1}$ is the constant observable, and ν_i is the sampling measure of the data used by the identifier. Denoting by $\Pi_{\mathcal{D}_i}$ the $L^2(\Omega_i, \nu_i)$ orthogonal projection onto $\mathcal{D}_i$, the closure bound of Assumption 1 is

$$\bar{r}_i = \sup_{\chi_i \in \Omega_i} \left\| (\mathcal{U}_i \Psi_i)(\chi_i) - (\Pi_{\mathcal{D}_i} \mathcal{U}_i \Psi_i)(\chi_i) \right\|, \tag{2.5}$$

the projection being applied componentwise to the p_i scalar observables collected in Ψ_i. When F_i is locally Lipschitz on Ω_i, and Ψ_i is continuous, $\mathcal{U}_i \Psi_i$ is bounded on Ω_i, so $\bar{r}_i < \infty$. Moreover, $\bar{r}_i \to 0$ as the dictionary is enriched toward a (control-)invariant subspace, with quantitative kernel-EDMD bounds available [12, 17, 30]. We do not assume $\bar{r}_i$ is known; it enters the guarantees only through the *certified* bound (3.4), and it is estimated empirically in Section 5, where its network mean stays near 0.6 cm after convergence. The novelty relative to single-system Koopman MPC is the explicit cross-coupling term $\sum_j W_{ij}^\star z_j$, which lifts the *interconnection* into the observable space and is what makes a distributed design and a small-gain certificate possible.

Remark 2.2 (Richness of the projection space). Definition (2.4) answers two questions that the notation $\operatorname{span} \Psi_i$ would leave open. First, the projection space *does* contain input-dependent and cross-subsystem observables; it has to, as the successor lift $z_i(k+1)$ genuinely depends on $u_i(k)$ and on $\{z_j(k)\}_{j \in \mathcal{N}_i}$, and a projection onto $\operatorname{span} \Psi_i$ alone could never reproduce that dependence. What (2.4) restricts is not the presence of these observables but their *degree*: $\mathcal{D}_i$ is the largest space in which the predictor is simultaneously affine in u_i and affine in each z_j. Second, this degree restriction is a design choice with a clear price and a clear payoff. Its payoff is that $\mathbb{P}_i$ in (4.2) remains a convex quadratic program and

that the coupling appears as constant matrices $W_{ij}^\star$ on which the small-gain certificate (4.15) can be imposed. Its price is that any genuinely input-bilinear or state-cross-product behavior is pushed into $\bar{r}_i$. When that price is too high, $\mathcal{D}_i$ can be enlarged by adjoining products such as $\{u_{i,l}\,\psi_{i,s}\}$ or $\{\psi_{i,s}\psi_{j,t}\}$, which keeps (2.3) *linear in the parameters*, so the identifier (3.1) and Theorem 3.3 apply verbatim, with q_i enlarged accordingly at the cost of the convexity of $\mathbb{P}_i$, which then requires the bilinear Koopman machinery of [26, 27] or an input transformation. The adequacy of the chosen degree is not left to faith: The residual diagnostic reported in Figure 5 measures $\bar{r}_i$ along the realized trajectories and therefore detects, at run time, a projection space that is too poor.

Remark 2.3. Equation (2.3) strictly generalizes the block-oriented description used in self-tuning regulators for interconnected Hammerstein systems [5, 6]: A Hammerstein model is recovered by choosing the dictionary to contain the basis of the static input nonlinearity and restricting $A_i^\star$ to the companion form of the linear block. The Koopman lifting removes the requirement that the nonlinearity act only at the input and that the dynamic block be linear in the original coordinates.

Collecting the unknowns, define the regressor and parameter matrix

$$\begin{aligned} \varphi_i(k) &\triangleq \operatorname{col}\big(z_i(k),\, u_i(k),\, \{z_j(k)\}_{j\in\mathcal{N}_i},\, 1\big) \in \mathbb{R}^{q_i}, \\ \Theta_i^\star &\triangleq \big[\, A_i^\star \;\; B_i^\star \;\; \{W_{ij}^\star\}_{j\in\mathcal{N}_i} \;\; e_i^\star \,\big]^\top \in \mathbb{R}^{q_i\times p_i}, \end{aligned} \tag{2.6}$$

with $q_i = p_i + m_i + \sum_{j\in\mathcal{N}_i} p_j + 1$ so that (2.3) reads $z_i(k+1)^\top = \varphi_i(k)^\top\Theta_i^\star + r_i(k)^\top$. The components of φ_i are exactly the basis functions spanning $\mathcal{D}_i$ in (2.4), evaluated along the trajectory; this is the sense in which the identifier of Section 3 estimates the projection (2.5) from data. To model aging and fouling, we allow slow drift of the lifted parameters.

Assumption 2. *The lifted parameters may be time-varying, $\Theta_i^\star = \Theta_i^\star(k)$, with $\big\|\Theta_i^\star(k) - \Theta_i^\star(k-1)\big\|_F \le \mu_i$ for all k.*

Assumption 3. *With the normalized regressor $\bar{\varphi}_i(k) \triangleq \varphi_i(k)/m_i(k)$, $m_i(k) \triangleq \sqrt{1 + \|\varphi_i(k)\|^2}$, there exist $\underline{\alpha}_i > 0$ and $S_i \in \mathbb{N}$ such that $\sum_{t=k+1}^{k+S_i} \bar{\varphi}_i(t)\bar{\varphi}_i(t)^\top \ge \underline{\alpha}_i I_{q_i}$ for all k.*

Normalization guarantees $\|\bar{\varphi}_i\| < 1$ and is standard in robust identification [2, 45]; Assumption 3 is the usual excitation requirement, here imposed on the *lifted* regressor, and is satisfied in the closed loop by a small dither [33, 56].

3. Adaptive Koopman identification

The design and its guarantees rest on three results that build directly on the preliminaries above and motivate one another. The first concerns the online identifier and shows that the recursive least-squares law keeps the covariance bounded and drives the parameter error into a small residual ball (Theorem 3.3). The second propagates this accuracy through the network and bounds the multistep prediction error (Theorem 4.3). The third closes the loop and proves input-to-state stability of the interconnected closed loop under a verifiable small-gain condition (Theorem 4.8), together with recursive feasibility (Proposition 4.11). We begin with the identifier.

Each subsystem maintains an estimate $\hat{\Theta}_i(k)$ of $\Theta_i^\star$ from the data stream $\{z_i, u_i, \{z_j\}_{j\in\mathcal{N}_i}\}$ by the normalized recursive least-squares law with forgetting factor $\lambda_i \in (0, 1)$:

$$\zeta_i(k) = z_i(k+1)/m_i(k), \qquad \bar{e}_i(k) = \zeta_i(k) - \hat{\Theta}_i(k-1)^\top\bar{\varphi}_i(k), \tag{3.1a}$$

$$K_i(k) = \frac{P_i(k-1)\bar{\varphi}_i(k)}{\lambda_i + \bar{\varphi}_i(k)^\top P_i(k-1)\bar{\varphi}_i(k)}, \tag{3.1b}$$

$$\hat{\Theta}_i(k) = \hat{\Theta}_i(k-1) + K_i(k)\,\bar{e}_i(k)^\top, \tag{3.1c}$$

$$P_i(k) = \Pi_{[\underline{p}_i, \bar{p}_i]}\left(\frac{1}{\lambda_i}\Big(P_i(k-1) - K_i(k)\bar{\varphi}_i(k)^\top P_i(k-1)\Big)\right), \tag{3.1d}$$

where $\Pi_{[\underline{p}_i, \bar{p}_i]}(\cdot)$ projects the (symmetric) covariance onto the spectral interval $[\underline{p}_i, \bar{p}_i]$ by clipping its eigenvalues, $0 < \underline{p}_i \le \bar{p}_i < \infty$. The current model matrices $\hat{A}_i(k), \hat{B}_i(k), \hat{W}_{ij}(k), \hat{e}_i(k)$ are read off from $\hat{\Theta}_i(k)$ through (2.6). This projection is the safeguard that prevents the covariance wind-up (estimator bursting) afflicting forgetting-factor RLS when excitation is temporarily lost in closed loop; it is the eigenvalue-clipping form of the classical constant-trace recursive estimator of self-tuning control [2, 45].

Lemma 3.1. *Let Assumption 3 hold, let $P_i(0) = p_{0,i}I$ with $\underline{p}_i \le p_{0,i} \le \bar{p}_i$, and let the projection interval be chosen so that*

$$\underline{p}_i \le 1 - \lambda_i, \qquad \bar{p}_i \;\ge\; (\lambda_i^{S_i - 1}\underline{\alpha}_i)^{-1}. \tag{3.2}$$

Then, the iterates of (3.1) *satisfy the following conditions:*

(i) $\underline{p}_i I \le P_i(k) \le \bar{p}_i I$ *for every* $k \ge 0$.

(ii) The lower clip in (3.1d) *is never active, and the upper clip is not active at any* $k \ge S_i$.

(iii) Consequently, whatever the clipping did during the initial transient $k < S_i$*, the covariance obeys the* unprojected *information recursion*

$$P_i(k)^{-1} = \lambda_i P_i(k-1)^{-1} + \bar{\varphi}_i(k)\bar{\varphi}_i(k)^\top \tag{3.3}$$

for every $k > S_i$.

Proof. See Appendix A. □

Remark 3.2 (What the projection does during the transient). Part (iii) of Lemma 3.1 is what licenses the use of (3.3) in the proof of Theorem 3.3, and it deserves emphasis because the closed form of the unprojected recursion is *not* available during the transient: While excitation is still being accumulated, the clipping may genuinely modify P_i, so the explicit solution (A.2) of the unclipped recursion cannot be invoked at those instants. The proof in Appendix A therefore avoids that closed form altogether. It uses only two facts that hold irrespective of how often the clipping acted: Eigenvalue clipping is *monotone* in the positive-semidefinite order, so it can only *increase* the information matrix P_i^{-1}, and the projection itself *enforces* $P_i(k) \in [\underline{p}_i I, \bar{p}_i I]$ at every instant. Chaining the first fact over one excitation window and combining it with Assumption 3 already places the *candidate* covariance inside the admissible interval for all $k \ge S_i$, so from that instant, the clipping has nothing left to do. The design conditions (3.2) are explicit and easy to satisfy: For the numerical study of Section 5, $\lambda_i = 0.97$ gives $1 - \lambda_i = 3 \cdot 10^{-2}$ against the implemented $\underline{p}_i = 10^{-2}$, and $\bar{p}_i = 2 \cdot 10^2$, so both conditions in (3.2) hold, and the identifier operates in the unclipped regime after the seeding phase.

Theorem 3.3. *Let Assumptions 1–3 hold, and let $\hat{\Theta}_i$ be generated by* (3.1) *with $\lambda_i \in (0, 1)$ and the projection interval of Lemma 3.1. Then, the parameter error $\tilde{\Theta}_i(k) \triangleq \hat{\Theta}_i(k) - \Theta_i^\star(k)$ satisfies, for all*

$k \geq S_i$,

$$\left\|\tilde{\Theta}_i(k)\right\|_F \leq \frac{\bar{p}_i}{\underline{p}_i}\,\lambda_i^{k-S_i}\,\left\|\tilde{\Theta}_i(S_i)\right\|_F + \frac{\bar{p}_i}{1-\lambda_i}\left(\bar{r}_i + \underline{p}_i^{-1}\mu_i\right). \tag{3.4}$$

In particular, $\limsup_{k\to\infty}\left\|\tilde{\Theta}_i(k)\right\|_F \leq \varepsilon_i \triangleq \bar{p}_i(\bar{r}_i + \underline{p}_i^{-1}\mu_i)/(1-\lambda_i)$*: The estimate converges exponentially* to the residual set $\{\Theta : \left\|\Theta - \Theta_i^\star(k)\right\|_F \leq \varepsilon_i\}$*, whose radius is proportional to the closure error* $\bar{r}_i$ *and to the drift* μ_i.

Proof. By Lemma 3.1(iii), the projection (3.1d) is inactive for $k > S_i$, so on that horizon, the covariance obeys the information form (3.3), which inverts (3.1d) without projection by the matrix inversion lemma. Combining (3.1b) with (3.3) gives the gain identity

$$P_i(k)^{-1}K_i(k) = P_i(k)^{-1}\frac{P_i(k-1)\bar{\varphi}_i(k)}{\lambda_i + \bar{\varphi}_i(k)^\top P_i(k-1)\bar{\varphi}_i(k)} = \bar{\varphi}_i(k), \tag{3.5}$$

where the last equality uses $P_i(k)^{-1}P_i(k-1) = \lambda_i I + \bar{\varphi}_i\bar{\varphi}_i^\top P_i(k-1)$ so that $P_i(k)^{-1}P_i(k-1)\bar{\varphi}_i = (\lambda_i + \bar{\varphi}_i^\top P_i(k-1)\bar{\varphi}_i)\bar{\varphi}_i$. Write the drift $\Delta_i(k) \triangleq \Theta_i^\star(k) - \Theta_i^\star(k-1)$ so that $\|\Delta_i(k)\|_F \leq \mu_i$ by Assumption 2. Using (2.3) as $\zeta_i(k) = \Theta_i^\star(k)^\top\bar{\varphi}_i(k) + \bar{r}_i(k)$ with $\bar{r}_i(k) \triangleq r_i(k)/m_i(k)$ and $\|\bar{r}_i(k)\| \leq \bar{r}_i$ (because $m_i \geq 1$), the a-priori error (3.1a) becomes

$$\bar{e}_i(k) = \Theta_i^\star(k)^\top\bar{\varphi}_i(k) + \bar{r}_i(k) - \hat{\Theta}_i(k-1)^\top\bar{\varphi}_i(k) = -\big(\tilde{\Theta}_i(k-1) - \Delta_i(k)\big)^\top\bar{\varphi}_i(k) + \bar{r}_i(k) \tag{3.6}$$

because $\hat{\Theta}_i(k-1) - \Theta_i^\star(k) = \tilde{\Theta}_i(k-1) - \Delta_i(k)$. Substituting into (3.1c) and subtracting $\Theta_i^\star(k)$ gives the error recursion

$$\tilde{\Theta}_i(k) = \big(I - K_i(k)\bar{\varphi}_i(k)^\top\big)\tilde{\Theta}_i(k-1) + K_i(k)\bar{r}_i(k)^\top - \big(I - K_i(k)\bar{\varphi}_i(k)^\top\big)\Delta_i(k), \tag{3.7}$$

and (3.3) yields $I - K_i\bar{\varphi}_i^\top = \lambda_i P_i(k)P_i(k-1)^{-1}$. Define the information state $G_i(k) \triangleq P_i(k)^{-1}\tilde{\Theta}_i(k)$. Premultiplying (3.7) by $P_i(k)^{-1}$ using (3.5) and the identity $P_i(k)^{-1}(I - K_i\bar{\varphi}_i^\top) = \lambda_i P_i(k-1)^{-1}$, one obtains the stable scalar-driven filter

$$G_i(k) = \lambda_i\, G_i(k-1) + \bar{\varphi}_i(k)\bar{r}_i(k)^\top - \lambda_i P_i(k-1)^{-1}\Delta_i(k). \tag{3.8}$$

Taking Frobenius norms and using $\|\bar{\varphi}_i\| < 1$, $\|\bar{r}_i\| \leq \bar{r}_i$, $P_i(k-1)^{-1} \leq \underline{p}_i^{-1}I$ (Lemma 3.1), and $\|\Delta_i\|_F \leq \mu_i$,

$$\|G_i(k)\|_F \leq \lambda_i\,\|G_i(k-1)\|_F + \bar{r}_i + \lambda_i\,\underline{p}_i^{-1}\mu_i. \tag{3.9}$$

Iterating this bound from $k = S_i$ and summing the geometric series gives, for $k \geq S_i$,

$$\begin{aligned}\|G_i(k)\|_F &\leq \lambda_i^{k-S_i}\,\|G_i(S_i)\|_F + \frac{1-\lambda_i^{k-S_i}}{1-\lambda_i}(\bar{r}_i + \underline{p}_i^{-1}\mu_i)\\ &\leq \lambda_i^{k-S_i}\,\|G_i(S_i)\|_F + \frac{\bar{r}_i + \underline{p}_i^{-1}\mu_i}{1-\lambda_i}.\end{aligned} \tag{3.10}$$

Returning to the parameter error, $\tilde{\Theta}_i(k) = P_i(k)G_i(k)$ with $\|P_i(k)\| \leq \bar{p}_i$ (Lemma 3.1), so $\left\|\tilde{\Theta}_i(k)\right\|_F \leq \bar{p}_i\,\|G_i(k)\|_F$. Likewise, $\|G_i(S_i)\|_F = \left\|P_i(S_i)^{-1}\tilde{\Theta}_i(S_i)\right\|_F \leq \underline{p}_i^{-1}\left\|\tilde{\Theta}_i(S_i)\right\|_F$. Combining the two displays with (3.10) yields exactly (3.4). The decaying factor $\lambda_i^{k-S_i} \to 0$ establishes exponential convergence to the residual set of radius $\varepsilon_i = \bar{p}_i(\bar{r}_i + \underline{p}_i^{-1}\mu_i)/(1-\lambda_i)$, which proves the lim sup claim. □

Remark 3.4. Theorem 3.3 isolates the two irreducible error sources of any finite-dictionary adaptive Koopman scheme: The closure error $\bar{r}_i$ and the drift μ_i. It is worth being precise about what the theorem does and does not claim. What decays exponentially is the *distance to the residual set* of radius ε_i, not the parameter error itself: Unless $\bar{r}_i = \mu_i = 0$, that is, unless the dictionary spans a control-invariant subspace, *and* the plant is exactly time-invariant in the lifted coordinates, $\tilde{\Theta}_i$ does not converge to zero, and $\varepsilon_i > 0$ is a genuine, irreducible floor. All downstream statements—the prediction bound (4.6) and the ISS estimate (4.18)—inherit this floor, which is why they are formulated with respect to a residual set rather than an equilibrium. The *rate* is nevertheless exponential with ratio λ_i, and this is what makes the recovery after a disturbance or a fault fast and quantifiable (Corollary 3.5). Both floors can be driven down, $\bar{r}_i$ by enriching the dictionary toward an invariant subspace [12, 17] and $\mu_i/(1-\lambda_i)$ by choosing λ_i closer to one when drift is slow. The bound is the lifted-space analog of the prediction-error convergence results that underpin self-tuning regulators [2, 4], and unlike a frozen EDMD model [24], it does not assume the data distribution is stationary.

3.1. Abrupt parameter changes: Jumps and dwell time

Assumption 2 models fouling and aging, which are slow by nature. A valve fault, a sensor recalibration, or a load switch is not: It changes the plant in one sampling period, and the increment $\left\|\Theta_i^\star(k) - \Theta_i^\star(k-1)\right\|_F$ then exceeds any small μ_i at that single instant. Theorem 3.3 therefore does *not* apply verbatim across such an event, and the numerical study of Section 5—which deliberately injects an instantaneous 50% change of an outlet coefficient—sits precisely in this regime. The following corollary covers it. It replaces the uniform bound of Assumption 2 by a piecewise-constant model with isolated jumps, which is the standard description of faults in adaptive and fault-tolerant control.

Corollary 3.5 (Piecewise-constant parameters with dwell time)**.** *Let Assumptions 1 and 3 hold, and replace Assumption 2 by the following: The lifted parameters are piecewise constant; that is, there is a strictly increasing sequence of jump instants $\mathcal{J}_i = \{k_i^1 < k_i^2 < \cdots\}$ with $\Theta_i^\star(k) = \Theta_i^\star(k_i^\ell)$ for $k \in [k_i^\ell, k_i^{\ell+1})$, jump sizes bounded by $\left\|\Theta_i^\star(k_i^\ell) - \Theta_i^\star(k_i^\ell - 1)\right\|_F \le \bar{\mu}_i$ and dwell time $k_i^{\ell+1} - k_i^\ell \ge T_i^{\mathrm{d}}$ for all ℓ. Then, for every ℓ and every $k \in [k_i^\ell, k_i^{\ell+1})$ with $k \ge S_i$,*

$$\left\|\tilde{\Theta}_i(k)\right\|_F \;\le\; \frac{\bar{p}_i}{\underline{p}_i}\,\lambda_i^{k-k_i^\ell}\Big(\left\|\tilde{\Theta}_i(k_i^\ell - 1)\right\|_F + \bar{\mu}_i\Big) \;+\; \frac{\bar{p}_i}{1-\lambda_i}\,\bar{r}_i. \tag{3.11}$$

Consequently:

(i) Each jump produces an excursion of the parameter error of size at most $(\bar{p}_i/\underline{p}_i)\,\bar{\mu}_i$, which is bounded *and decays geometrically at rate λ_i.*

(ii) The error returns to within ϵ of the drift-free residual set $\{\left\|\tilde{\Theta}_i\right\|_F \le \bar{p}_i\bar{r}_i/(1-\lambda_i)\}$ after at most

$$T_i^{\mathrm{rec}}(\epsilon) \;=\; \left\lceil \frac{\ln\!\big(\underline{p}_i\,\epsilon \,/\, (\bar{p}_i\,(\varepsilon_i^- + \bar{\mu}_i))\big)}{\ln \lambda_i} \right\rceil \;\; \textit{samples}, \qquad \varepsilon_i^- \triangleq \frac{\bar{p}_i\bar{r}_i}{1-\lambda_i}, \tag{3.12}$$

so if the dwell time satisfies $T_i^{\mathrm{d}} \ge T_i^{\mathrm{rec}}(\epsilon)$, the identifier has recovered before the next jump.

(iii) Averaged over a jump cycle, the sequence $\Theta_i^\star$ satisfies Assumption 2 with the mean *drift rate $\mu_i^{\mathrm{avg}} = \bar{\mu}_i/T_i^{\mathrm{d}}$, and* (3.4) *holds in the averaged sense with μ_i replaced by μ_i^{avg}.*

Proof. On the open interval $(k_i^\ell, k_i^{\ell+1})$, the parameters are constant, so $\Delta_i(k) = 0$, and (3.8) reduces to $G_i(k) = \lambda_i G_i(k-1) + \bar{\varphi}_i(k)\bar{r}_i(k)^\top$; iterating from k_i^ℓ and using $\|\bar{\varphi}_i\| < 1$ gives $\|G_i(k)\|_F \le \lambda_i^{k-k_i^\ell} \left\|G_i(k_i^\ell)\right\|_F + \bar{r}_i/(1-\lambda_i)$. At the jump instant itself, $\Delta_i(k_i^\ell)$ is the single nonzero increment of the Frobenius norm at most $\bar{\mu}_i$, so $\left\|\tilde{\Theta}_i(k_i^\ell)\right\|_F \le \left\|\tilde{\Theta}_i(k_i^\ell - 1)\right\|_F + \bar{\mu}_i$ by the triangle inequality applied to $\tilde{\Theta}_i(k_i^\ell) = \hat{\Theta}_i(k_i^\ell) - \Theta_i^\star(k_i^\ell - 1) - \Delta_i(k_i^\ell)$. Converting between G_i and $\tilde{\Theta}_i$ through $\underline{p}_i^{-1}$ and $\bar{p}_i$ exactly as at the end of the proof of Theorem 3.3 yields (3.11). Claim (i) follows by taking $k = k_i^\ell$, claim (ii) by solving $(\bar{p}_i/\underline{p}_i)\lambda_i^T(\varepsilon_i^- + \bar{\mu}_i) \le \epsilon$ for T, and claim (iii) follows by noting that a jump of size $\bar{\mu}_i$ every T_i^{d} samples has the same total variation over a cycle as a uniform drift of rate $\bar{\mu}_i/T_i^{\mathrm{d}}$, which is the quantity that enters (3.10) through the geometric sum. □

Remark 3.6 (Reading of Corollary 3.5). Three consequences are worth recording. First, the certificate survives the jump: Nothing diverges because the excursion is bounded by the jump size amplified by the fixed conditioning factor $\bar{p}_i/\underline{p}_i$ of the covariance interval, and no reinitialization of the identifier is required. Second, the recovery is geometric with the *same* ratio λ_i that governs the nominal transient, so (3.12) turns the forgetting factor into a directly interpretable design parameter: With $\lambda_i = 0.97$, the error is divided by e every $-1/\ln\lambda_i \approx 33$ samples, and Section 5 measures a recovery of about 60 samples, that is, roughly two such time constants. Third, the closed loop inherits this behavior without a separate argument: By Theorem 4.8, the loop is ISS with respect to $d_i \propto \varepsilon_i$, and (3.11) says that a jump makes d_i momentarily larger but bounded and then geometrically decaying. ISS therefore turns it into a bounded tracking excursion followed by a return to the nominal residual set, and Proposition 4.11, whose argument is instant-by-instant, is unaffected. The tracking cost of the jump is quantified in Table 4.

3.2. Update frequency, computational load, and measurement noise

The law (3.1) is executed at every sampling instant, which may seem inconsistent with the premise that the parameters drift slowly, and it is written for noise-free lifted data, which no real plant provides. This subsection settles both points quantitatively.

3.2.1. Intermittent updates

Suppose the estimate is refreshed only on the update grid $\mathcal{T}_u = \{0, T_u, 2T_u, \dots\}$ with $T_u \in \mathbb{N}$ and is held constant in between. The regression itself remains one-step; that is, at $k \in \mathcal{T}_u$, the pair $(\varphi_i(k), z_i(k+1))$ is used. Two things change: Excitation is now required on windows of S_i *updates*; that is, $T_u S_i$ samples, and the parameters may move by up to $T_u\mu_i$ between two consecutive updates.

Corollary 3.7 (Periodic updating)**.** *Let Assumptions 1–2 hold, let Assumption 3 hold along the update grid $\mathcal{T}_u$, and let* (3.1) *be executed only for $k \in \mathcal{T}_u$. Then, for the nth update instant $k = nT_u \ge S_i T_u$,*

$$\left\|\tilde{\Theta}_i(k)\right\|_F \le \frac{\bar{p}_i}{\underline{p}_i}\,\lambda_i^{n-S_i}\left\|\tilde{\Theta}_i(S_iT_u)\right\|_F + \frac{\bar{p}_i}{1-\lambda_i}\Big(\bar{r}_i + \underline{p}_i^{-1}\,T_u\mu_i\Big), \tag{3.13}$$

and between updates, the error grows by at most μ_i per sample. Hence, the asymptotic radius is $\varepsilon_i(T_u) = \bar{p}_i(\bar{r}_i + \underline{p}_i^{-1}T_u\mu_i)/(1-\lambda_i) + T_u\mu_i$, which grows affinely *in T_u, whereas the decay rate per* sample *degrades from λ_i to λ_i^{1/T_u}, and the identifier cost falls from $O(q_i^2 p_i)$ to $O(q_i^2 p_i)/T_u$ per sampling instant.*

Proof. Re-index (3.8) by the update counter n. Between two updates, the estimate is frozen while $\Theta_i^\star$ moves by at most $T_u\mu_i$ in Frobenius norm (Assumption 2 applied T_u times), so the drift increment seen by the nth update is $\left\|\Delta_i^{(n)}\right\|_F \le T_u\mu_i$. The closure residual is unchanged because the regression stays one-step. Applying the proof of Theorem 3.3 verbatim in the index n, with μ_i replaced by $T_u\mu_i$, gives (3.13); the additional $T_u\mu_i$ in $\varepsilon_i(T_u)$ accounts for the intersample hold, and the per-sample rate follows from $\lambda_i^n = \lambda_i^{k/T_u}$. □

The design message is that T_u buys computation linearly and costs accuracy linearly, with no threshold effect to exploit, and the second penalty—the T_u-fold slower *recovery*—is the one that matters after a fault, not the steady-state radius. Because the identifier is by far the cheaper of the two online computations, $O(q_i^2 p_i)$ against $O(H^2 m_i)$ per quadratic program and measured at under 5% of the per-subsystem solve time in Section 5, there is little to gain: Slowing the estimator by an order of magnitude removes about 4% of the per-step cost while inflating the tracking error by more than 45% and tripling the fault recovery time (Table 5). Intermittent updating is therefore advisable only when the identifier shares a processor with a much heavier task or when T_u is used deliberately as a low-pass filter against noise, which is the subject of the following.

3.2.2. Measurement noise

Let the levels be measured with bounded error $x_i^{\mathrm{m}}(k) = x_i(k) + n_i(k)$ with $\|n_i(k)\| \le \bar{n}_i$, and let the identifier and the controller use the corrupted lift $z_i^{\mathrm{m}}(k) = \Psi_i(x_i^{\mathrm{m}}(k))$. By Definition 2.1, the lifting is Lipschitz, so $\left\|z_i^{\mathrm{m}} - z_i\right\| \le L_{\psi,i}\bar{n}_i \triangleq \epsilon_i^n$; the noise therefore corrupts *both* the regressor and the regression target, which is an errors-in-variables situation and not a benign additive output disturbance.

Corollary 3.8 (Bounded measurement noise)**.** *Let Assumptions 1–3 hold for the noise-free data, and let* $\bar{\theta}_i \triangleq \sup_k \left\|\Theta_i^\star(k)\right\|_F$. *Then, the noisy data satisfy* (2.3) *with* r_i *replaced by an effective residual* r_i^n *obeying*

$$\bar{r}_i^{\,n} \;\triangleq\; \sup_k \left\|r_i^n(k)\right\| \;\le\; \bar{r}_i + \epsilon_i^n + \bar{\theta}_i\Big(\epsilon_i^n + \sum_{j\in\mathcal{N}_i}\epsilon_j^n\Big), \tag{3.14}$$

and Theorem 3.3 and all its consequences hold verbatim with $\bar{r}_i$ *replaced by* $\bar{r}_i^n$. *In particular, the asymptotic radius inflates to* $\varepsilon_i^n = \bar{p}_i(\bar{r}_i^n + \underline{p}_i^{-1}\mu_i)/(1-\lambda_i)$, *which is affine in the noise level* $\bar{n}_i$ *with slope* $\bar{p}_i(1+\bar{\theta}_i)L_{\psi,i}/(1-\lambda_i)$ *plus the neighbors' contribution.*

Proof. Write $\varphi_i^{\mathrm{m}} = \varphi_i + \tilde{\varphi}_i$ and $z_i^{\mathrm{m}}(k+1) = z_i(k+1) + \Delta_i^n(k+1)$, where $\left\|\Delta_i^n\right\| \le \epsilon_i^n$, and because the input is known exactly, and the constant regressor is exact, $\|\tilde{\varphi}_i\| \le \epsilon_i^n + \sum_{j\in\mathcal{N}_i}\epsilon_j^n$. Substituting in (2.3), $z_i^{\mathrm{m}}(k+1)^\top = \varphi_i^{\mathrm{m}}(k)^\top\Theta_i^\star + \left(r_i(k) + \Delta_i^n(k+1) - \Theta_i^{\star\top}\tilde{\varphi}_i(k)\right)^\top$, and the bracket is r_i^n bounded as in (3.14) by the triangle and submultiplicative inequalities. The remainder of the proof of Theorem 3.3 uses only the bound on the residual, not its statistics. □

Corollary 3.8 is deliberately worst-case, and its honest reading is that the noise contribution $\bar{\theta}_i\,\|\tilde{\varphi}_i\|$ is a *bias*, not a zero-mean term that averages out: Because $\tilde{\varphi}_i$ is correlated with φ_i, more data do not remove it. Three practical remedies exist and trade off differently.

(i) *Raise* λ_i. The radius ε_i^n is proportional to $(1-\lambda_i)^{-1}$ only through the *sum* $\bar{r}_i^n + \underline{p}_i^{-1}\mu_i$, whereas the variance of the estimate falls as $(1-\lambda_i)$; increasing λ_i therefore attenuates the noise-driven fluctuation at the price of a slower response to drift and to faults, which is the classical tracking-versus-noise compromise already recorded in Remark 5.2(iii) and quantified in Table 4.

(ii) *Dead zone.* Freezing the update whenever the measured output residual falls below a threshold $\varrho_i \propto \bar{n}_i$ removes the noise-driven random walk of $\hat{\Theta}_i$; the price is that the residual set of Theorem 3.3 is enlarged by the dead-zone radius because errors smaller than ϱ_i are no longer corrected. The threshold must therefore be matched to the noise level, and Table 6 shows how sharp that requirement is: A dead zone set at half the noise standard deviation suppresses a quarter of the updates at no cost in accuracy, whereas one set at twice the standard deviation suppresses more than 80% of them and, by then blocking legitimate adaptation, degrades tracking beyond the level of the unprotected estimator.

(iii) *Bias compensation.* Instrumental-variable or bias-compensated recursive schemes remove the errors-in-variables term at the price of requiring a noise-variance estimate and of losing the finite-time worst-case bound; they are the natural route when $\bar{n}_i$ is large and is known.

Finally, the intermittent update of Corollary 3.7 and the dead zone of item (ii) act on the same quantity from opposite directions: The first reduces the number of noisy updates by a fixed factor irrespective of the data, the second reduces it adaptively, only when the information content is below the noise floor. The numerical comparison is reported in Section 5.

4. Adaptive Koopman distributed predictive control

At each instant k, subsystem i uses its current estimate to predict its lifted trajectory over a horizon H under a candidate input sequence $\mathbf{u}_i = \{u_i(\ell \mid k)\}_{\ell=0}^{H-1}$,

$$\hat{z}_i(\ell+1 \mid k) = \hat{A}_i \hat{z}_i(\ell \mid k) + \hat{B}_i u_i(\ell \mid k) + \sum_{j \in \mathcal{N}_i} \hat{W}_{ij}\, \zeta_j(\ell \mid k) + \hat{e}_i, \quad \hat{z}_i(0 \mid k) = z_i(k), \tag{4.1}$$

with output $\hat{x}_i(\ell \mid k) = C_i \hat{z}_i(\ell \mid k)$, where $\zeta_j(\ell \mid k)$ is the lifted trajectory broadcast by neighbor j at the previous step (one communication round per instant, the assumed-trajectory paradigm of distributed MPC [42,43]). We collect the data of the local problem at time k in the pair $(\hat{\Theta}_i(k), \boldsymbol{\zeta}_i(k))$, where $\boldsymbol{\zeta}_i(k) \triangleq \{\zeta_j(\cdot \mid k)\}_{j \in \mathcal{N}_i}$. Subsystem i solves the finite-horizon optimal control problem $\mathbb{P}_i(z_i(k); \hat{\Theta}_i(k), \boldsymbol{\zeta}_i(k))$:

$$\min_{\mathbf{u}_i} \sum_{\ell=0}^{H-1} \Big(\big\|\hat{x}_i(\ell \mid k) - x_i^{\mathrm{r}}\big\|_{Q_i}^2 + \big\|u_i(\ell \mid k) - u_i^{\mathrm{r}}\big\|_{R_i}^2 \Big) + \big\|\hat{x}_i(H \mid k) - x_i^{\mathrm{r}}\big\|_{P_i^f}^2 \tag{4.2}$$

subject to (4.1), $u_i(\ell \mid k) \in \mathbb{U}_i$, and the terminal constraint $\hat{x}_i(H \mid k) \in \mathbb{X}_i^f$, with $Q_i, R_i, P_i^f \succ 0$ and u_i^{r} the steady input of Assumption 4 that holds x_i^{r}. The receding-horizon law applies the first move, $u_i(k) = u_i^\star(0 \mid k)$. The cost is written in deviation form so that it vanishes at the target equilibrium $(x_i^{\mathrm{r}}, u_i^{\mathrm{r}})$, which makes the optimal value minimal there (Lemma 4.6).

Remark 4.1 (Arguments of the local value function). The optimal value of (4.2) is *not* a function of x_i alone. It depends on the current estimate $\hat{\Theta}_i(k)$ through the matrices $(\hat{A}_i, \hat{B}_i, \hat{W}_{ij}, \hat{e}_i)$ that define the predictor (4.1) and on the assumed neighbor trajectories $\boldsymbol{\zeta}_i(k)$, which enter (4.1) as a known affine forcing term. We therefore write the optimal value as

$$V_i(x_i; \hat{\Theta}_i, \boldsymbol{\zeta}_i) \triangleq \text{optimal value of } \mathbb{P}_i(\Psi_i(x_i); \hat{\Theta}_i, \boldsymbol{\zeta}_i), \qquad V_i^k(x_i) \triangleq V_i(x_i; \hat{\Theta}_i(k), \boldsymbol{\zeta}_i(k)), \tag{4.3}$$

and we use the abbreviation V_i^k whenever the data are those in force at time k. This is more than notation: Because $\hat{\Theta}_i$ and $\boldsymbol{\zeta}_i$ are updated between k and $k+1$, the function V_i^k is genuinely *time-varying*, and a

Lyapunov argument based on it must account for the change of the function itself and not only for the motion of its argument. Lemma 4.6(iii) supplies the required sensitivity bound, and Lemma 4.7 bounds the per-step change of the data; the proof of Theorem 4.8 then carries both terms explicitly.

Remark 4.2. Because (4.1) is affine in $(\hat{z}_i, \mathbf{u}_i)$, and the stage cost is convex quadratic, $\mathbb{P}_i$ is a convex quadratic program of dimension Hm_i, solvable to global optimality in real time; condensing the prediction yields the dense Hessian $2(S_u^\top \bar{Q} S_u + \bar{R})$ used in Section 5. The optimizations are decoupled given the neighbor trajectories and hence solved in parallel. This is the distributed, certainty-equivalence counterpart of the centralized Koopman MPC of [24].

Assumption 4. *The reference is an admissible interior equilibrium: There is a steady input $u_i^{\mathrm{r}} \in \operatorname{int} \mathbb{U}_i$ for which x_i^{r} is an equilibrium of the nominal lifted predictor with lifted reference $z_i^{\mathrm{r}} \triangleq \Psi_i(x_i^{\mathrm{r}})$. Moreover, for each i, there exist a terminal weight $P_i^f \succ 0$, a terminal set $\mathbb{X}_i^f$ with $x_i^{\mathrm{r}} \in \operatorname{int} \mathbb{X}_i^f$, and a local terminal feedback κ_i^f with $\kappa_i^f(x_i^{\mathrm{r}}) = u_i^{\mathrm{r}}$ such that for the nominal local predictor ($\bar{r}_i = 0$, $\varepsilon_i = 0$ and frozen neighbor state), $\mathbb{X}_i^f$ is positively invariant under κ_i^f with $\kappa_i^f(x_i) \in \mathbb{U}_i$, and the terminal cost decrease $\left\|x_i^+ - x_i^{\mathrm{r}}\right\|_{P_i^f}^2 - \left\|x_i - x_i^{\mathrm{r}}\right\|_{P_i^f}^2 \le -\left\|x_i - x_i^{\mathrm{r}}\right\|_{Q_i}^2 - \left\|\kappa_i^f(x_i) - u_i^{\mathrm{r}}\right\|_{R_i}^2$ holds on $\mathbb{X}_i^f$.*

Assumption 4 is the standard quasi-infinite-horizon terminal construction for setpoint tracking [57–59], here imposed on the lifted linear model so that P_i^f and $\mathbb{X}_i^f$ follow from a local linear quadratic regulator/linear matrix inequalities (LQR/LMI) computation. Taking $\mathbb{X}_i^f$ as a small sublevel set of $\left\|x_i - x_i^{\mathrm{r}}\right\|_{P_i^f}^2$ places x_i^{r} in its interior; the interior steady input u_i^{r} makes the input constraint slack on a neighborhood of the reference, which is what renders V_i a smooth quadratic there (Lemma 4.6).

Assumption 5. *For every estimate $\hat{\Theta}_i$ in the projection set the pair $(\hat{A}_i, \hat{B}_i)$ is stabilizable, and a local gain K_i renders the prestabilized local predictor $A_{K,i} \triangleq \hat{A}_i + \hat{B}_i K_i$ Schur, with $\left\|A_{K,i}\right\| \le a_i < 1$ in a suitable (weighted) norm. The terminal set $\mathbb{X}_i^f$ of Assumption 4 is* robustly *positively invariant for the prestabilized predictor against the neighbor-coupling disturbance set $\mathbb{W}_i \triangleq \{\sum_{j \in \mathcal{N}_i} \hat{W}_{ij}(z_j - z_j^{\mathrm{r}}) : \left\|z_j - z_j^{\mathrm{r}}\right\| \le \bar{z}_j^f\}$, where $\bar{z}_j^f$ bounds the neighbor deviation in the terminal region.*

Assumption 5 closes two gaps of certainty-equivalence Koopman MPC. First, the predictor is operated in the *prestabilized* form $u_i(\ell \mid k) = u_i^{\mathrm{r}} + K_i(\hat{z}_i(\ell \mid k) - z_i^{\mathrm{r}}) + v_i(\ell \mid k)$, where v_i is the new decision variable; this offset form gives $v_i = 0$ at the reference, consistent with $\kappa_i^f(x_i^{\mathrm{r}}) = u_i^{\mathrm{r}}$ (Assumption 4), and $\mathbb{P}_i$ is solved equivalently in $\mathbf{v}_i$ through this affine map. The error recursion below then propagates through the contractive $A_{K,i}$ instead of the possibly non-Schur open loop $\hat{A}_i$ [58]. Second, the terminal set is invariant against the neighbors, so a single terminal triple stays valid for every estimate in the projection set. The condensation of Section 4 is unchanged by this affine substitution, and the empirical spectral radius of the prestabilized lifted network, reported in Section 5, stays at $0.54 < 1$. The assumption is a strong (common) robust prestabilizability requirement: One gain K_i and one terminal triple must serve every estimate in the projection set. This is reasonable when that set is kept small by slow drift and rich excitation, and it can be enforced offline by a robust-invariance or LMI computation over a polytopic over-approximation of the set [35, 57]. In Section 5 a fixed LQR triple is enough because the identified models stay in a neighborhood where it remains valid. A constructive synthesis of robustly valid terminal ingredients for the full estimate set is left open; the practical consequences of this and of the other standing assumptions are collected in Section 4.2.

We first quantify how the estimation and closure errors propagate through the predictor. Let $\delta_i(\ell \mid k) \triangleq \hat{z}_i(\ell \mid k) - z_i(k+\ell)$ denote the ℓ-step lifted prediction error along the realized input, and collect the stacked errors in

$$\delta(\ell) \triangleq \operatorname{col}\big(\|\delta_1(\ell \mid k)\|, \dots, \|\delta_N(\ell \mid k)\|\big).$$

Theorem 4.3. *Let Assumptions 1, 2, and 5 hold, and suppose $\left\|\tilde{\Theta}_i(k)\right\|_F \le \varepsilon_i$ with the lifted state and input bounded by $\bar{z}, \bar{u}$ and the neighbor assumed-trajectory mismatch bounded by $\left\|\zeta_j(\ell \mid k) - \hat{z}_j^{\text{actual}}(\ell \mid k)\right\| \le \bar{\eta}_j$ ($\bar{\eta}_j$ models communication delay or packet loss and vanishes, $\bar{\eta}_j = 0$, under the synchronous one-round exchange of Section 4). Then, with $a_i \triangleq \left\|A_{K,i}\right\| < 1$, $w_{ij} \triangleq \left\|\hat{W}_{ij}\right\|$ and the horizon-dependent offset*

$$c_i(\ell) \triangleq \big((1 + |\mathcal{N}_i|)\bar{z} + \bar{u} + 1\big)\varepsilon_i + \bar{r}_i + \ell\,\mu_i\bar{\varphi}_i^{\max} + \sum_{j\in\mathcal{N}_i} w_{ij}\bar{\eta}_j, \tag{4.4}$$

where the term $\ell\mu_i\bar{\varphi}_i^{\max}$ accounts for the parameter drift accumulated over the ℓ steps already elapsed in the prediction (Assumption 2), and $\bar{\varphi}_i^{\max}$ bounds the regressor, the prediction error obeys the comparison inequality

$$\delta(\ell+1) \le M\,\delta(\ell) + c(\ell), \qquad \delta(0) = 0 \tag{4.5}$$

componentwise, where $c(\ell) \triangleq \operatorname{col}(c_1(\ell), \dots, c_N(\ell))$, and $M \triangleq D_a + \Gamma_w$ with $D_a = \operatorname{diag}(a_1, \dots, a_N)$ and $[\Gamma_w]_{ij} = w_{ij}$ for $j \in \mathcal{N}_i$, zero otherwise. Consequently, for every horizon $\ell \le H$,

$$\delta(\ell) \le \sum_{t=0}^{\ell-1} M^t\, c(\ell-1-t) \;\le\; \Big(\textstyle\sum_{t=0}^{\ell-1} M^t\Big) c(\ell-1), \tag{4.6}$$

and if $\rho(M) < 1$, then $\delta(\ell) \le (I - M)^{-1} c(H-1)$ uniformly in $\ell \le H$.

Proof. Subtract the prestabilized predictor (4.1), operated as $u_i = u_i^{\mathrm{r}} + K_i(\hat{z}_i - z_i^{\mathrm{r}}) + v_i$ (Assumption 5), from the true lifted dynamics (2.3) along the realized input. Writing $\tilde{A}_i \triangleq \hat{A}_i - A_i^\star(k)$ and, likewise, $\tilde{B}_i, \tilde{W}_{ij}, \tilde{e}_i$ — all referred to the parameter value $\Theta_i^\star(k)$ *at the current instant* — the homogeneous part propagates through $A_{K,i} = \hat{A}_i + \hat{B}_i K_i$:

$$\begin{aligned}
\delta_i(\ell+1 \mid k) = {} & A_{K,i}\,\delta_i(\ell \mid k) + \tilde{A}_i z_i(k+\ell) + \tilde{B}_i u_i(k+\ell) + \tilde{e}_i - r_i(k+\ell) \\
& - \sum_{s=0}^{\ell-1} \Delta\Theta_i^\star(k+s)^\top \varphi_i(k+s) \\
& + \sum_{j\in\mathcal{N}_i} \Big(\hat{W}_{ij}\big(\zeta_j(\ell \mid k) - z_j(k+\ell)\big) + \tilde{W}_{ij} z_j(k+\ell)\Big),
\end{aligned} \tag{4.7}$$

where $\Delta\Theta_i^\star(k+s) \triangleq \Theta_i^\star(k+s+1) - \Theta_i^\star(k+s)$ is the drift accumulated *after* the current instant. By Assumption 2, $\sum_{s=0}^{\ell-1} \left\|\Delta\Theta_i^\star(k+s)\right\|_F \le \ell\mu_i$, and $\|\varphi_i\| \le \bar{\varphi}_i^{\max}$, so this term is bounded by $\ell\mu_i\bar{\varphi}_i^{\max}$. The neighbor mismatch splits as $\zeta_j(\ell \mid k) - z_j(k+\ell) = (\zeta_j - \hat{z}_j^{\text{actual}}) + (\hat{z}_j^{\text{actual}} - z_j)$, with the first bracket bounded by $\bar{\eta}_j$ and the second equal to $\delta_j(\ell \mid k)$. Taking norms in (4.7) and using $\left\|A_{K,i}\right\| \le a_i$, $\left\|z_j\right\| \le \bar{z}$, $\|u_i\| \le \bar{u}$, $\|r_i\| \le \bar{r}_i$, the drift bound, and $\left\|\tilde{A}_i\right\| + \left\|\tilde{B}_i\right\| + \sum_j \left\|\tilde{W}_{ij}\right\| + \|\tilde{e}_i\| \le (1 + |\mathcal{N}_i|)\left\|\tilde{\Theta}_i\right\|_F \le (1 + |\mathcal{N}_i|)\varepsilon_i$ (each block norm is dominated by $\left\|\tilde{\Theta}_i\right\|_F$),

$$\|\delta_i(\ell+1 \mid k)\| \le a_i\,\|\delta_i(\ell \mid k)\| + \sum_{j\in\mathcal{N}_i} w_{ij}\left\|\delta_j(\ell \mid k)\right\| + c_i(\ell) \tag{4.8}$$

with $c_i(\ell)$ as in (4.4), which is (4.5) in vector form. Because M has non-negative entries, and $\delta(0) = 0$, induction on ℓ gives the first inequality in (4.6): If $\delta(\ell) \le \sum_{t=0}^{\ell-1} M^t c(\ell-1-t)$, then $\delta(\ell+1) \le M\delta(\ell) + c(\ell) \le \sum_{t=1}^{\ell} M^t c(\ell-t) + c(\ell) = \sum_{t=0}^{\ell} M^t c(\ell-t)$, the first step using monotonicity of $v \mapsto Mv$ on the non-negative orthant. The second inequality follows because $\ell \mapsto c_i(\ell)$ is nondecreasing, so $c(\ell-1-t) \le c(\ell-1)$ for every $t \ge 0$. Finally, if $\rho(M) < 1$, the Neumann series converges, and $\sum_{t=0}^{\ell-1} M^t \le (I-M)^{-1}$, which gives the uniform bound. □

Remark 4.4 (The two drift contributions do not overlap). The offset (4.4) contains the drift twice, once inside ε_i through Theorem 3.3 and once explicitly as $\ell\mu_i\bar{\varphi}_i^{\max}$, and it is worth verifying that this is not a double count. The two terms measure different quantities, separated by the instant k at which the prediction starts. The first, $\varepsilon_i \ge \left\|\hat{\Theta}_i(k) - \Theta_i^\star(k)\right\|_F$, is the *estimation lag at time k*: How far the identifier is from the parameter that is in force when the prediction is initialized. The second is the *variation of the true parameter after time k*, $\sum_{s=0}^{\ell-1} \left\|\Theta_i^\star(k+s+1) - \Theta_i^\star(k+s)\right\|_F \le \ell\mu_i$, which no estimator running up to time k can know. Formally, the separation is enforced in (4.7) by referring every tilde quantity to $\Theta_i^\star(k)$ and telescoping the remainder; the two contributions are therefore disjoint, and their sum is the correct bound, not a conservative duplication. Making the offset ℓ-dependent also sharpens the result where it matters. The uniform choice $H\mu_i\bar{\varphi}_i^{\max}$ charges the full horizon to every prediction step, whereas (4.4) charges only the steps already elapsed; over the horizon, this replaces $H\sum_t M^t$ by $\sum_t (H-1-t)M^t$, a reduction of roughly one half for small $\rho(M)$. The one-step error, which is what enters the closed-loop analysis of Theorem 4.8, tightens by the full factor H: It is governed by $c_i(1)$, in which the drift enters as μ_i and not as $H\mu_i$.

The condition $\rho(M) < 1$ is a small-gain condition on the prediction error: The identified predictors must be individually contractive after the terminal feedback, and the lifted coupling gains w_{ij} must be small relative to that contraction. It is the structural sibling of the closed-loop condition of Theorem 4.8, and (4.6) shows that the predictor error is, term by term, of order $\varepsilon_i + \bar{r}_i$ and hence vanishes with the dictionary error and the drift through Theorem 3.3.

4.1. Making the update aware of the multistep error

Theorem 4.3 bounds the multistep error, but the identifier (3.1) minimizes a *one-step* lifting residual, so it does not act on that bound directly. It is therefore natural to ask how the multistep error could be penalized inside the update. Equation (4.6) makes the question precise because it factorizes the H-step error into two independent handles,

$$\delta(H) \;\le\; \mathcal{A}_H\, c(H-1), \qquad \mathcal{A}_H \triangleq \sum_{t=0}^{H-1} M^t, \qquad \|\mathcal{A}_H\| \le \frac{1-\|M\|^H}{1-\|M\|} \;\; (\|M\| < 1). \tag{4.9}$$

The *residual level c* is what the one-step criterion already minimizes. The *amplification* $\mathcal{A}_H$, by contrast, depends only on the contraction rates a_i and the coupling norms $w_{ij} = \left\|\hat{W}_{ij}\right\|$; a one-step criterion has no grip on it whatsoever, and it is exactly the quantity that converts an acceptable one-step fit into an unacceptable H-step forecast when the network is strongly coupled or weakly contractive. A multistep-aware update must therefore trade the first handle against the second.

The cheapest way to do so that preserves both convexity and exact recursiveness is to charge the amplification directly in the estimation criterion. Replace the weighted least-squares objective

underlying (3.1) by

$$J_{i,k}(\Theta) = \sum_{t=1}^{k} \lambda_i^{k-t} \left\|\zeta_i(t) - \Theta^\top \bar{\varphi}_i(t)\right\|^2 + (1-\lambda_i)\, \nu_i \, \left\|D_i^{1/2}\Theta\right\|_F^2, \qquad D_i \triangleq \operatorname{diag}\left(I_{p_i},\, 0_{m_i},\, \{I_{p_j}\}_{j\in\mathcal{N}_i},\, 0\right), \tag{4.10}$$

where the selector D_i picks out exactly the rows of Θ that carry $\hat{A}_i$ and $\{\hat{W}_{ij}\}$, that is, the blocks that build M and leaves $\hat{B}_i$ and the offset $\hat{e}_i$ untouched so that the steady-state gain and the equilibrium of the predictor are not biased. The criterion (4.10) is still a convex quadratic in Θ, so its minimizer is still obtained recursively: (3.1a)–(3.1d) are unchanged, and the parameter update (3.1c) acquires one extra term,

$$\hat{\Theta}_i(k) = \hat{\Theta}_i(k-1) + K_i(k)\bar{e}_i(k)^\top - (1-\lambda_i)\, \nu_i\, P_i(k)\, D_i\, \hat{\Theta}_i(k-1), \tag{4.11}$$

which is the leakage (or σ-modification) form familiar from robust adaptive control [3]. Its cost is one matrix product, $O(q_i^2 p_i)$, the same order as the update it augments, so the real-time budget is unaffected.

Corollary 4.5 (Amplification–bias tradeoff)**.** *Under the hypotheses of Theorem 3.3, let $\hat{\Theta}_i$ be generated by* (3.1) *with* (3.1c) *replaced by* (4.11)*, and let $\bar{\theta}_i \triangleq \sup_k \left\|\Theta_i^\star(k)\right\|_F$. If $\nu_i < \bar{p}_i^{-1}$, the parameter error still converges exponentially to a residual set, now of radius*

$$\varepsilon_i^\nu = \frac{\varepsilon_i + \bar{p}_i \nu_i \bar{\theta}_i}{1 - \bar{p}_i \nu_i} \;\geq\; \varepsilon_i, \tag{4.12}$$

and the penalized blocks are contracted so that $\|M\|$, and with it, the amplification $\|\mathcal{A}_H\|$ of (4.9) *is nonincreasing in ν_i. The resulting H-step bound $\|\mathcal{A}_H(\nu_i)\|\, \|c(H-1;\nu_i)\|$ is thus the product of a nonincreasing and a nondecreasing function of ν_i and admits, in general, an interior minimizer.*

Proof. Premultiplying (4.11) by $P_i(k)^{-1}$ and repeating the steps that led to (3.8) adds the single term $-(1-\lambda_i)\nu_i D_i \hat{\Theta}_i(k-1)$ to the right-hand side of the Frobenius norm at most $(1-\lambda_i)\nu_i(\bar{\theta}_i + \left\|\tilde{\Theta}_i(k-1)\right\|_F)$ because $\|D_i\|_F \leq 1$. Carrying this through (3.10) adds $\bar{p}_i\nu_i(\bar{\theta}_i + \varepsilon_i^\nu)$ to the asymptotic radius; that is, $\varepsilon_i^\nu = \varepsilon_i + \bar{p}_i\nu_i(\bar{\theta}_i + \varepsilon_i^\nu)$, which solves to (4.12) for $\nu_i < \bar{p}_i^{-1}$. Monotonicity of the penalized blocks in ν_i follows from the fact that (4.10) is a ridge-regularized least-squares criterion in the corresponding rows. □

Two remarks temper the enthusiasm this construction might raise, and both are borne out by the numerical study of Section 5. First, the mechanism pays only when $\mathcal{A}_H$ is actually large. On the tank benchmark $\rho(M) \approx 0.5$ so that $\|\mathcal{A}_H\| \approx 2$, and the H-step error is dominated by the residual level, not by the amplification; Table 7 accordingly shows a small improvement of the one-step fit at $\nu_i = 0.005$ and no net gain on the eight-step error, with a clear degradation once ν_i grows. The construction is therefore a tool for strongly coupled or weakly contractive networks, where $\rho(M)$ approaches one, and $\|\mathcal{A}_H\|$ grows without bound, and not a free improvement. Second, the alternatives are less attractive than they look. Fitting a *separate* predictor per horizon (direct multihorizon EDMD) keeps the problem convex and recursive, but the resulting maps are regressed on $\varphi_i(k)$ alone and cannot express the dependence on the future inputs $u_i(k+1), \ldots, u_i(k+\ell-1)$, so they are unusable inside a receding-horizon program. Weighting the genuine ℓ-step residuals $\|z_i(k+\ell) - \hat{z}_i(\ell \mid k)\|^2$ makes the criterion nonconvex in Θ because $\hat{z}_i(\ell \mid k)$ involves $\hat{A}_i^\ell$, and the exact recursive solution is lost — one then falls back on the damped Gauss–Newton or multistep recursive schemes of [37], which retain a per-sample cost but no finite-time

bound of the form (3.4). Equation (4.9) is what makes the compromise of (4.10) attractive: It buys a grip on the amplification while keeping every property that Theorem 3.3 rests on.

We can now state the main closed-loop result. Following standard MPC practice, we use the local optimal value (4.3) as a local ISS Lyapunov function. The following regularity properties hold under Assumptions 1, 4, and 5.

Lemma 4.6. *There exist constants $0 < q_i \le \bar{c}_i$, $L_i > 0$, $\eta_i > 0$, and $L_i^{\Theta}, L_i^{\zeta}, L_i^{(2)} \ge 0$ such that writing $s_i \triangleq \left\|x_i - x_i^{\mathrm{r}}\right\|$ and $\mathcal{N}_i \triangleq \{x_i : s_i \le \eta_i\}$, the following hold.*

(i) (Sandwich.) *On the feasible set, $q_i s_i^2 \le V_i(x_i; \hat{\Theta}_i, \boldsymbol{\zeta}_i) \le \bar{c}_i s_i^2$ with $q_i = \lambda_{\min}(Q_i)$ uniformly over the admissible data.*

(ii) (Local smoothness.) *The radius $\eta_i > 0$ can be chosen so that neither the terminal constraint nor the input constraints of $\mathbb{P}_i$ are active on $\mathcal{N}_i$. For fixed data, $V_i(\cdot; \hat{\Theta}_i, \boldsymbol{\zeta}_i)$ is then* exactly *the quadratic $V_i(x_i) = (1/2)(x_i - x_i^{\mathrm{r}})^{\top}\Pi_i(x_i - x_i^{\mathrm{r}})$ on $\mathcal{N}_i$ with $\Pi_i \succ 0$ the constant condensed value Hessian; hence, V_i is twice continuously differentiable there, with $V_i(x_i^{\mathrm{r}}) = 0$, $\nabla V_i(x_i^{\mathrm{r}}) = 0$ and $\left\|\nabla^2 V_i\right\| = \|\Pi_i\| \le L_i$, the constant L_i being computable in closed form from the quadratic program data.*

(iii) (Sensitivity to the problem data.) *For every $x_i \in \mathcal{N}_i$ and every pair of admissible data $(\Theta, \boldsymbol{\zeta})$, $(\Theta', \boldsymbol{\zeta}')$,*

$$\left|V_i(x_i; \Theta', \boldsymbol{\zeta}') - V_i(x_i; \Theta, \boldsymbol{\zeta})\right| \le \Big(L_i^{\Theta}\,\|\Theta' - \Theta\|_F + L_i^{\zeta}\sum_{j\in\mathcal{N}_i}\left\|\zeta_j' - \zeta_j\right\|\Big)s_i + L_i^{(2)}\Big(\|\Theta' - \Theta\|_F^2 + \sum_{j\in\mathcal{N}_i}\left\|\zeta_j' - \zeta_j\right\|^2\Big). \tag{4.13}$$

Proof. See Appendix B. □

Lemma 4.7 (Per-step variation of the local data)**.** *Under the hypotheses of Theorem 3.3, the estimate produced by* (3.1) *varies by at most*

$$\varsigma_i \triangleq \limsup_{k\to\infty}\left\|\hat{\Theta}_i(k) - \hat{\Theta}_i(k-1)\right\|_F \le \frac{\bar{p}_i}{\lambda_i}(\varepsilon_i + \mu_i + \bar{r}_i) \tag{4.14}$$

per sampling instant; in particular, $\varsigma_i = O(\bar{r}_i + \mu_i)$, of the same order as the estimation error itself. Moreover, the neighbor data satisfy $\left\|\zeta_j(\cdot \mid k+1) - \zeta_j(\cdot \mid k)\right\| \le L_{\psi,j}(s_j(k) + s_j(k+1)) + \bar{\eta}_j$.

Proof. By (3.1c), $\hat{\Theta}_i(k) - \hat{\Theta}_i(k-1) = K_i(k)\bar{e}_i(k)^{\top}$. From (3.1b), $\|K_i(k)\| \le \|P_i(k-1)\|\,\|\bar{\varphi}_i(k)\|/\lambda_i \le \bar{p}_i/\lambda_i$, using $\|\bar{\varphi}_i\| < 1$, $\|P_i\| \le \bar{p}_i$ (Lemma 3.1) and the fact that the denominator of (3.1b) is at least λ_i. From (3.6) and $\|\bar{\varphi}_i\| < 1$, $\|\bar{e}_i(k)\| \le \left\|\tilde{\Theta}_i(k-1)\right\|_F + \|\Delta_i(k)\|_F + \bar{r}_i$, which is asymptotically at most $\varepsilon_i + \mu_i + \bar{r}_i$ by Theorem 3.3 and Assumption 2. The neighbor estimate follows from the Lipschitz property of Ψ_j (Definition 2.1) applied to the broadcast trajectories, both of which lie within $\bar{\eta}_j$ of the neighbor's realized lift. □

Theorem 4.8. *Let Assumptions 1–5 hold, let $\mathbb{P}_i$ be feasible at $k = 0$ for every i, and define the gain matrix $\Gamma = [\gamma_{ij}] \in \mathbb{R}^{N\times N}$ with γ_{ij} constructed* explicitly *in* (C.10) *of Appendix C from the identified coupling $\hat{W}_{ij}$; the curvature and dictionary constants $L_i, L_{\psi,j}$; the value-function sandwich constants $\bar{c}_i, q_j$; and free Young parameters $\varepsilon_{ij} > 0$ (and $\gamma_{ij} = 0$ for $j \notin \mathcal{N}_i$). If*

$$\rho(\Gamma) < 1, \tag{4.15}$$

let $v \succ 0$ satisfy $\Gamma v \prec v$, and define the network Lyapunov function and the operating region

$$V^k(x) \triangleq \max_{i\in\mathcal{V}} \frac{V_i^k(x_i)}{v_i}, \qquad \Omega_v \triangleq \{x : \ V^k(x) \le c^\star\}, \qquad c^\star \triangleq \min_{i\in\mathcal{V}} \frac{q_i\eta_i^2}{v_i}, \tag{4.16}$$

with q_i, η_i from Lemma 4.6. Then, the interconnected closed loop under the law (4.2) *is ISS on Ω_v with respect to $d \triangleq \mathrm{col}(d_1, \dots, d_N)$,*

$$d_i \triangleq \varepsilon_i + \bar{r}_i + \mu_i\bar{\varphi}_i^{\max} + \varsigma_i, \tag{4.17}$$

which gathers the estimation, closure, one-step drift, and value function variation terms of (4.4) *and* (4.14)*: There exist $\beta \in \mathcal{KL}$, $\gamma \in \mathcal{K}_\infty$, and $d^\star > 0$ such that for every $x(0) \in \Omega_v$ and every disturbance with $\|d\|_\infty \le d^\star$, the state remains in Ω_v, and*

$$\|x(k) - x^{\mathrm{r}}\| \le \beta(\|x(0) - x^{\mathrm{r}}\|, k) + \gamma(\|d\|_\infty), \quad \forall k \ge 0. \tag{4.18}$$

Combined with Theorem 3.3 ($\varepsilon_i \to \bar{p}_i(\bar{r}_i + \underline{p}_i^{-1}\mu_i)/(1-\lambda_i)$), the asymptotic tracking error satisfies $\limsup_k \left\|x_i(k) - x_i^{\mathrm{r}}\right\| \le \gamma_i'(\max_j(\bar{r}_j + \mu_j))$ for some $\gamma_i' \in \mathcal{K}_\infty$ and can therefore be rendered arbitrarily small by enriching the dictionaries under slow drift.

Proof. See Appendix C. □

Remark 4.9 (Regional, not global). Theorem 4.8 is a *regional* statement, and the region Ω_v in (4.16) is exactly the set on which its ingredients are available. The reason is Lemma 4.6(ii): The curvature bound L_i, and with it the explicit gains (C.10), are obtained from the fact that V_i coincides with a single quadratic where no constraint is active, which is guaranteed only on $\mathcal{N}_i$. Outside $\mathcal{N}_i$, the value function of a strictly convex multiparametric quadratic program is still a continuous, convex, and piecewise quadratic [60] and hence locally Lipschitz, but it is not differentiable across the boundaries between critical regions, and its curvature is only piecewise defined. Two routes are then available, and both cost something. One can repeat the argument of Appendix C with the Lipschitz constant of V_i on a compact subset of the feasible set in place of L_i; the bilinear terms in (C.4) become $O(\|e\|)$ rather than $O(s_i\|e\|)$, so Young's inequality yields $\mathcal{K}$-function gains instead of the quadratic ones. The small-gain test becomes the cyclic condition of [46, 47] and the explicit formula (C.10) is lost. Alternatively, one can enlarge $\mathcal{N}_i$ by design by choosing $\mathbb{U}_i$ and $\mathbb{X}_i^f$ so that the unconstrained critical region covers the intended operating envelope, which is what is done in Section 5, where the pumps saturate only during the largest setpoint transients. We state the result regionally because that is what is proved with computable gains; the region is explicit, it is not infinitesimal, and it is verifiable from the same quadratic program data that produce L_i.

Remark 4.10. The small-gain condition (4.15) is verifiable from the *identified* models because Γ is assembled from $\left\|\hat{A}_i\right\|$, $\left\|\hat{W}_{ij}\right\|$ and the design weights; it is the same condition that governs the prediction error in Theorem 4.3. This is the structural advantage of lifting the interconnection into the observable space: The coupling becomes a set of constant matrices $\hat{W}_{ij}$ on which a quantitative, distributed stability certificate can be imposed, something not available to monolithic Koopman MPC [24, 41] or single-system adaptive Koopman schemes [38, 39].

Proposition 4.11. *Suppose the only state constraint of $\mathbb{P}_i$ is the terminal constraint $\hat{x}_i(H \mid k) \in \mathbb{X}_i^f$ (the input constraint $\mathbb{U}_i$ being a box), and let Assumptions 4–5 hold with the terminal set tightened to $\mathbb{X}_i^f \ominus \mathcal{B}_i$,*

where $\mathcal{B}_i$ is the ball of radius

$$\bar{\rho}_i \triangleq \|C_i\| \, \frac{1-a_i^H}{1-a_i}\Big(\kappa_i \bar{d}^\star + \sum_{j\in\mathcal{N}_i} \big\|\hat{W}_{ij}\big\| \, \bar{\eta}_j\Big), \qquad \bar{d}^\star \triangleq \max_i c_i(H-1), \tag{4.19}$$

that is, the worst-case prediction error accumulated through the contractive $A_{K,i}$ ($a_i < 1$). If $\bar{d}^\star$ and $\bar{\eta}_j$ are small enough that the tightened terminal set is nonempty, and the prestabilizing move keeps the shifted candidate in $\mathbb{U}_i$, then the feasibility of $\mathbb{P}_i$ at k implies feasibility at $k+1$ for all i.

Proof. See Appendix D. □

4.2. The standing assumptions: What can be relaxed and at what price

The guarantees above rest on five standing assumptions. Several of them are idealizations that a real plant will satisfy only approximately, and it is therefore legitimate to ask which of them can be weakened, how far, and what happens to the certificates when they are. This subsection answers those three questions in turn; Table 1 summarizes the outcome. The recurring theme is that the assumptions are not equally fragile: Two of them are essentially structural and cheap, two degrade gracefully, and one — the common prestabilizability of Assumption 5 — is the genuinely restrictive one.

Table 1. Standing assumptions, admissible relaxations, and their effect on the closed-loop certificates.

Assumption	Idealization	Practical relaxation	Effect on the certificate
1 Closure	Finite projection error, value unknown.	None needed; quantify $\bar{r}_i$ by kernel EDMD or a posteriori.	Qualitative ISS retained; quantitative bound needs $\bar{r}_i$.
2 Drift	Uniformly slow variation.	Piecewise constant with dwell time $T^{\mathrm{d}} \geq T^{\mathrm{rec}}$ (Corollary 3.5).	Bounded excursion, geometric recovery at rate λ_i.
3 Excitation	Uniform window excitation.	Excitation on average; dither; dead zone when uninformative.	Boundedness always retained; exponential convergence lost.
4 Terminal	Exact terminal triple.	Long horizon without terminal set; terminal equality.	Decrease factor $\alpha(H) < 1$; threshold on H.
5 Prestab.	One gain valid for all estimates.	Polytopic LMI design; online re-synthesis; scheduled family	Small-gain margin monitored online; certificate suspended if lost.

Assumption 1 (bounded closure residual). This is the mildest of the five assumptions. It does *not* require the dictionary to span a Koopman-invariant subspace; it requires only that the projection error (2.5) be finite on the compact operating set, which is automatic whenever F_i is continuous, and Ψ_i is bounded, as recorded after Assumption 1. What is genuinely not available is the *value* of $\bar{r}_i$: Without it, (4.18) certifies that the tracking error is bounded by a $\mathcal{K}_\infty$ function of the closure error but does

not say how large that bound is. The relaxation is therefore not of the assumption but of the claim: A quantitative certificate requires either kernel-EDMD error estimates [13, 30], which demand additional regularity of F_i, or the a posteriori diagnostic of Figure 5, which certifies $\bar{r}_i$ along the trajectories actually visited and nowhere else. We regard the second as the honest engineering answer and use it as such.

Assumption 2 (slow drift). Already relaxed in Corollary 3.5 to piecewise-constant parameters with isolated jumps and a dwell time — the model that actually describes faults. The certificate survives with a bounded excursion and a geometric recovery whose time constant (3.12) is explicit. The degree of relaxation is quantified by the dwell time: What may *not* be relaxed is that jumps be separated by more than T_i^{rec}, as a plant that jumps faster than the identifier can track is, by construction, not identifiable online. Between these two regimes lies fast continuous drift, for which the residual radius $\varepsilon_i \propto \mu_i/(1-\lambda_i)$ simply grows: The certificate does not fail, it becomes weak, and it announces the fact through ε_i.

Assumption 3 (persistency of excitation). This is the assumption most often violated in practice because a well-regulated loop is by definition poorly exciting. Its role should be located precisely. It is *not* needed for the covariance to stay bounded: Lemma 3.1(i) holds by construction of the projection, whatever the data, and this is exactly why the eigenvalue clipping is part of the design rather than a numerical convenience. Without excitation, what is lost is Lemma 3.1(ii)–(iii) and with them, the exponential contraction of (3.4): The estimate remains bounded but needs no longer approach the residual set, so ε_i can no longer be computed from $\bar{r}_i$ and μ_i. Theorem 4.8 still applies, as it takes ε_i as an input, but the conclusion weakens from "tracking error governed by dictionary richness" to "tracking error governed by whatever accuracy the estimator happens to retain". Practical relaxations are the standard ones: Excitation on average rather than on every window; a small dither, as used here; or switching the adaptation off through a dead zone when the regressor is uninformative, which, as Section 3.2 notes, is the same device that protects against noise.

Assumption 4 (terminal ingredients). This assumption is standard and standardly relaxable. The terminal set and weight may be replaced by a sufficiently long horizon without terminal constraint at the price of replacing the exact decrease (C.1) by one with a horizon-dependent factor $\alpha(H) \to 1$ so that Theorem 4.8 holds for H larger than a computable threshold. This effect can also be achieved by a terminal equality constraint, which shrinks the feasible set. Because here, the terminal triple is computed from the *lifted linear* predictor by an LQR calculation, the assumption costs little in practice; its residual difficulty is inherited from Assumption 5, namely that one triple should serve all estimates.

Assumption 5 (common robust prestabilizability). This is the restrictive one, and we say so plainly. It demands a single gain K_i and a single terminal triple valid for *every* estimate in the projection set, and it is the assumption most likely to be violated when the estimate moves quickly, for instance, immediately after a fault. Three relaxations are available in increasing order of implementation cost: Restricting the projection set to a polytope over which a common gain exists, computed offline by an LMI [35, 57]; resynthesizing $(K_i, P_i^f, \mathbb{X}_i^f)$ online whenever the estimate leaves a certified validity region, which turns the scheme into a switched design and requires a dwell-time argument; or scheduling a finite family of triples over a partition of the projection set. What happens when it is violated is detectable rather than silent: $\rho(M)$ and $\rho(\Gamma)$ are computed from the identified matrices at run time, so

a transient loss of the small-gain margin is observed as it happens (Figure 5, right), and the natural fallback—freezing the model, shortening the horizon, or reverting to the last certified estimate—can be triggered on that signal. During such a window, the multistep bound (4.6) and the terminal invariance of Proposition 4.11 are suspended; stability is not automatically lost, but it is no longer certified, and this is the correct statement to make.

The overall scheme is summarized in Algorithm 1.

Algorithm 1 Adaptive Koopman Distributed MPC (AK-DMPC), per subsystem i

1: **Initialize** $\hat{\Theta}_i(0)$ (offline EDMD seed), $P_i(0) = p_{0,i}I$, terminal ingredients $(P_i^f, \mathbb{X}_i^f, \kappa_i^f)$, projection interval $[\underline{p}_i, \bar{p}_i]$ satisfying (3.2).
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: measure $x_i(k)$; form $z_i(k) = \Psi_i(x_i(k))$; receive $\{\zeta_j\}_{j\in\mathcal{N}_i}$ from neighbors.
4: read $\hat{A}_i, \hat{B}_i, \hat{W}_{ij}, \hat{e}_i$ from $\hat{\Theta}_i(k)$; solve the QP $\mathbb{P}_i(z_i(k); \hat{\Theta}_i(k), \boldsymbol{\zeta}_i(k))$ in (4.2).
5: apply $u_i(k) = u_i^\star(0 \mid k)$; broadcast the predicted lifted trajectory $\{\hat{z}_i(\ell \mid k)\}_{\ell=1}^H$.
6: measure $x_i(k+1)$; update $\hat{\Theta}_i$ and P_i by (3.1).
7: **end for**

5. Simulation results

The proposed AK-DMPC approach is now utilized and validated on the considered system, a network of three interconnected nonlinear tanks. This benchmark exercises every ingredient of the design at once: The dynamics are nonlinear (square-root outflows), the subsystems are interconnected (coupling pipes between the tanks), the parameters are uncertain and drifting (an abrupt valve fault), and the inputs are constrained (pump saturation). It therefore provides a faithful and reproducible test of the online identifier of Section 3 and of the distributed predictive controller and its certificates in Section 4, and it is the setting on which all claims of the paper are confirmed.

Concretely, the considered system is a network of $N = 3$ interconnected nonlinear tanks connected in a line (1–2–3), a large-scale extension of the quadruple-tank process [52] that has recently served as a Koopman-MPC benchmark [53]. Each tank obeys a Torricelli balance with base-coupling pipes

$$A\dot{h}_i = u_i - a_i\sqrt{2gh_i} - \sum_{j\in\mathcal{N}_i} c_{ij}\,\mathrm{sgn}(h_i - h_j)\sqrt{2g|h_i - h_j|}, \tag{5.1}$$

with cross-section $A = 50\ \mathrm{cm}^2$, $g = 981\ \mathrm{cm/s}^2$, outlet coefficients $a = (3.8, 3.6, 4.0)$, pipe coefficients $c_{12} = 2.0$, $c_{23} = 2.2$, input saturation $u_i \in [0, 2200]\ \mathrm{cm}^3/\mathrm{s}$, and sampling period $T_s = 1$ s (twenty Euler substeps). The square-root outflow and the coupling make (5.1) genuinely nonlinear and interconnected; the controller is given no analytic model.

Each subsystem uses a state-inclusive dictionary of dimension $p_i = 4$,

$$\Psi_i(h) = \left[\, h,\ \sqrt{h+\delta},\ h^2,\ e^{-(h-30)^2/(2\cdot 18^2)} \,\right]^\top, \qquad \delta = 10^{-3}$$

so that $C_i = [1, 0, 0, 0]$. The identifier (3.1) uses $\lambda_i = 0.97$, constant-trace level $8q_i$, and eigenvalue bounds $[\underline{p}_i, \bar{p}_i] = [10^{-2}, 2\cdot 10^2]$, which satisfy the design conditions (3.2) because $1 - \lambda_i = 3 \cdot 10^{-2} \geq \underline{p}_i$. The predictive controller uses $H = 8$, $Q_i = 1$, $R_i = 2\cdot 10^{-6}$, and terminal weight $P_i^f = 5$. A 400 s

open-loop excitation phase (identical across all controllers) seeds the models; closed-loop tracking then runs for 700 s over a sequence of step references, and at $t = 350$ s, an abrupt valve fault increases a_2 by 50% (fouling). A shared input dither of standard deviation 9 cm^3/s sustains excitation, realizing the closed-loop persistency of Assumption 3 [33]. We compare AK-DMPC against three baselines: (B1) a *frozen* Koopman MPC (the seed model kept fixed, that is, the centralized scheme of [24] applied per subsystem); (B2) a *decentralized linear* MPC (a global least-squares first-order model per tank, ignoring the coupling); and (B3) an adaptive *Hammerstein self-tuning regulator* in the tradition of [5, 6], with the per-subsystem input–output model $y_i(k+1) = \theta_i^\top [\, y_i(k),\ \tilde{u}_i,\ \tilde{u}_i^2,\ \{y_j(k)\}_{j\in\mathcal{N}_i},\ 1\,]$ ($\tilde{u}_i = u_i/\bar{u}$) identified online by the same forgetting-factor RLS and closed with a generalized-minimum-variance move (control penalty $\rho_u = 2\cdot 10^{-6}$). Baseline B3 is the genuine block-oriented self-tuning competitor that AK-DMPC generalizes.

Figure 1 shows the level trajectories of all four controllers. Before the fault, every adaptive scheme tracks, but the decentralized linear controller carries persistent offsets because it ignores both the nonlinearity and the coupling. After the fault, the frozen Koopman controller develops a large steady-state bias on the faulted tank 2, whereas AK-DMPC and the Hammerstein regulator re-identify online and recover. Figure 2 shows the inputs; Figure 3 the one-step prediction error, whose abrupt rise and reconvergence after the fault is the signature of Theorem 3.3 (post-fault 0.62 cm for AK-DMPC against 1.46 cm frozen); and Figure 4 the per-tank RMSE.

Table 2 reports the single-run RMSE and Table 3 the Monte Carlo statistics over eight seeds. AK-DMPC attains a full-run mean RMSE of 1.07 cm, 28% below the frozen Koopman controller (1.48 cm) and 58% below the decentralized linear controller (2.55 cm); on the faulted tank, it lowers the post-fault RMSE from 3.41 cm (frozen) to 2.09 cm, and it shows the smallest cross-seed variance among the Koopman schemes (1.35 ± 0.20 versus 2.10 ± 0.49). The adaptive Hammerstein regulator (B3) is the strongest baseline—unsurprisingly, as a one-step minimum-variance regulator is near-optimal for a first-order plant—and is marginally more accurate on raw tracking (0.99 ± 0.28). It attains this, however, only through aggressive near-deadbeat action: Its input total variation is 3.1× that of AK-DMPC (Table 3), it cannot incorporate input constraints, and it carries *no* closed-loop stability or recursive-feasibility guarantee—a limitation inherited from classical self-tuning control [1, 5]. AK-DMPC thus matches the best block-oriented self-tuning regulator in accuracy while adding constraint handling, threefold smoother control, and the certificates of Section 4.

Table 2. Single-run network tracking RMSE (cm), mean over the three tanks.

Controller	Pre-fault	Post-fault	Full run
Decentralized linear MPC (B2)	2.59	2.48	2.55
Frozen Koopman MPC (B1) [24]	1.27	1.62	1.48
Hammerstein self-tuning regulator (B3) [6]	0.91	0.79	0.87
AK-DMPC (proposed)	**0.68**	1.34	1.07

Table 3. Monte Carlo results over eight seeds: Full-run network RMSE (mean ± std, cm) and mean per-tank input total variation (10^3 cm^3/s, control activity; lower is smoother).

Controller	RMSE (mean ± std)	input TV
Decentralized linear MPC (B2)	2.56 ± 0.07	15.4
Frozen Koopman MPC (B1)	2.10 ± 0.49	16.0
Hammerstein self-tuning regulator (B3)	0.99 ± 0.28	48.1
AK-DMPC (proposed)	1.35 ± 0.20	**15.7**

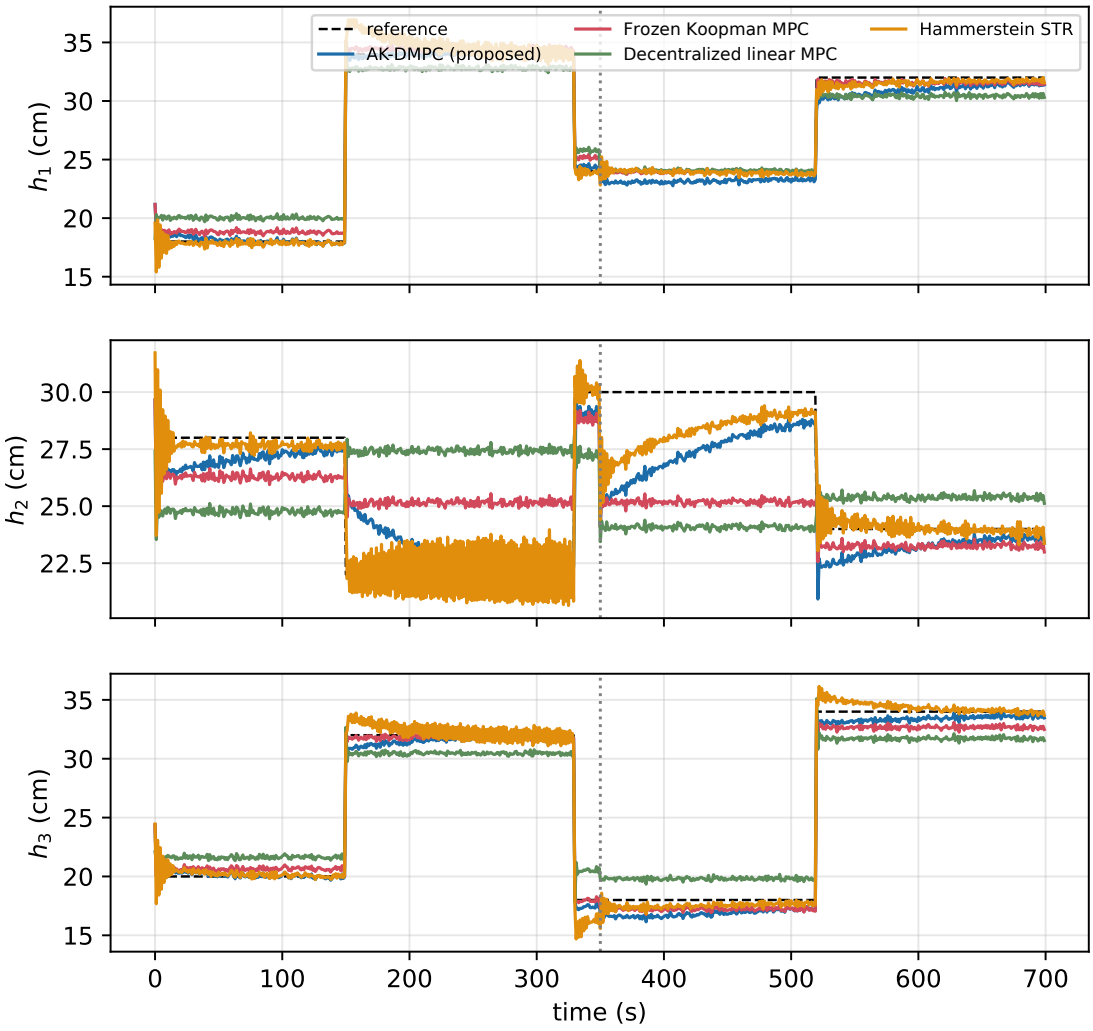

Figure 1. Level trajectories of the three tanks under the four controllers; the dashed line is the reference. The valve fault occurs at t = 350 s on tank 2.

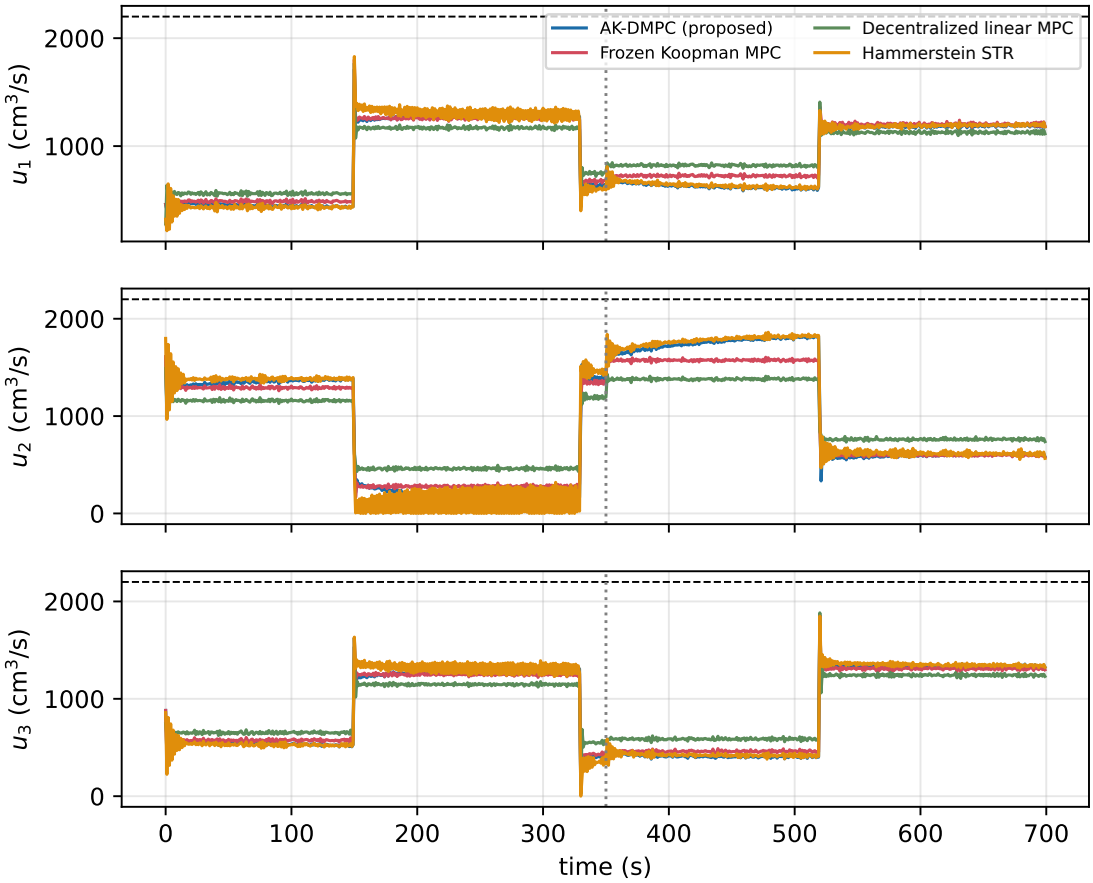

Figure 2. Control inputs; the dashed line is the saturation bound $\bar{u}$.

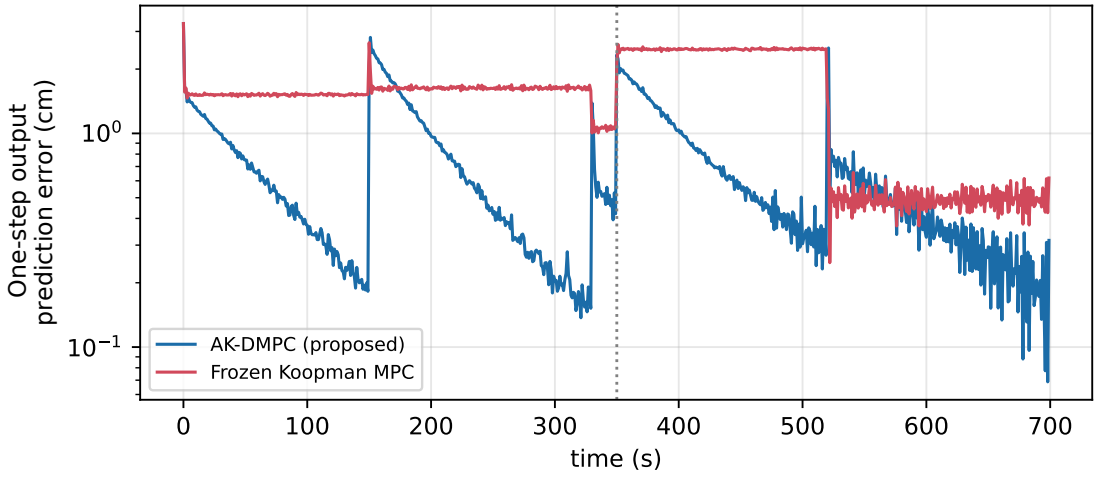

Figure 3. One-step output prediction error (log scale). After the fault, the frozen predictor degrades and remains inaccurate, whereas the adaptive predictor reconverges, as predicted by Theorem 3.3.

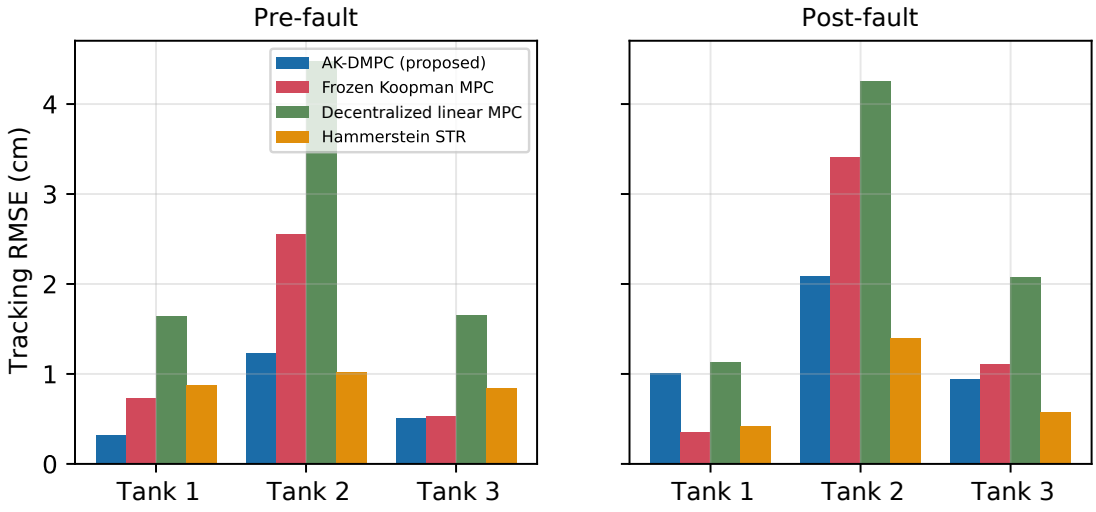

Figure 4. Per-tank tracking RMSE before and after the valve fault.

The proposed scheme also certifies its own premises online (Figure 5). The empirical closure residual $|C_i r_i|$ remains bounded with a network mean of 0.60 cm after convergence. We stress that this is an *a posteriori*, trajectory-based diagnostic of $\bar{r}_i$ on the tested trajectories, not a global certificate of the supremum (2.5) over the whole operating set; an a priori guarantee would require additional regularity (e.g., Lipschitz F_i together with kernel-EDMD error estimates [13, 30]). The spectral radius of the prestabilized lifted-network matrix stays at $\rho(A_{\mathrm{cl}}) \approx 0.54 < 1$ throughout, so the small-gain premise of Theorem 4.8 is verified from the identified models at run time (Assumption 5). Each local quadratic program is solved in 0.82 ms (condensed box-QP, L-BFGS-B), well within the 1 s period and trivially parallel across subsystems, confirming the real-time distributed claim of Remark 4.2.

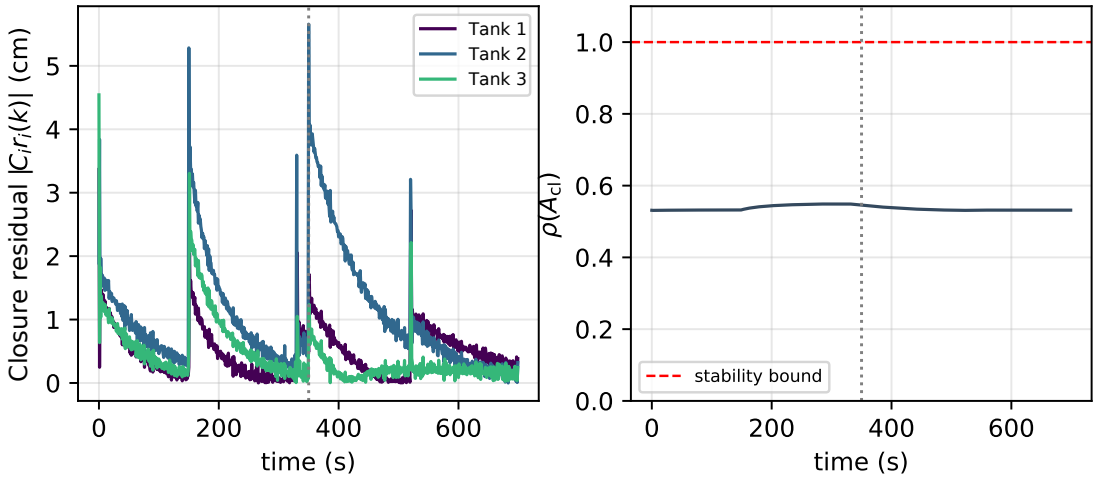

Figure 5. Online certificates: bounded closure residual and prestabilized lifted-network spectral radius. Left: The output closure residual $|C_i r_i|$ stays bounded (network mean 0.60 cm), empirically substantiating Assumption 1. Right: The spectral radius $\rho(A_{\mathrm{cl}}) \approx 0.54 < 1$ of the prestabilized lifted-network matrix verifies the small-gain premise of Theorem 4.8 from the identified models.

5.1. Behavior under the abrupt parameter jump

The valve fault injected at $t = 350$ s is an instantaneous 50% change and therefore falls under Corollary 3.5 rather than under Assumption 2. Table 4 isolates its effect. The upper block varies the forgetting factor at the nominal fault size, and the lower block varies the fault size at the nominal $\lambda_i = 0.97$; all entries are means over four seeds, and the recovery time is the number of samples after the fault at which the one-step output prediction error falls back below 1 cm and stays there.

Three observations follow, and each matches a claim of Corollary 3.5. First, the excursion is always bounded and always transient: No run diverges, and the identifier recovers without reinitialization, which is claim (i). Second, the recovery time grows with the size of the jump, from 19 samples for a +20% fault to 106 samples for a +100% fault, and at the nominal fault, it is about 60 samples, that is, roughly two time constants $-1/\ln \lambda_i \approx 33$ of the geometric decay in (3.11). The bound (3.12) is nevertheless very conservative in absolute terms because it charges the full covariance conditioning $\bar{p}_i/\underline{p}_i = 2 \cdot 10^4$ to the excursion, whereas the realized covariance is far better conditioned along the excited trajectory. Third—and this is the practically useful finding—the performance is almost insensitive to λ_i over the whole range $[0.90, 0.99]$, with the full-run RMSE varying by less than 2%. The reason is exactly claim (iii): Between faults, the parameters are *constant*, so $\mu_i = 0$, and the term $\mu_i/(1 - \lambda_i)$ that would normally penalize a large λ_i vanishes. The tracking-versus-noise compromise of λ_i therefore has little bite against *jumps*; it reappears, as Corollary 3.8 predicts, only when the data are noisy, which is the subject of Table 6.

Table 4. Behavior under the abrupt valve fault, means over four seeds. Upper block: Effect of the forgetting factor at the nominal +50% fault. Lower block: Effect of the fault magnitude at $\lambda_i = 0.97$. Recovery time is the number of samples after the fault at which the one-step output prediction error returns below 1 cm and remains there.

Setting	Pre-fault RMSE (cm)	Post-fault RMSE (cm)	Recovery (samples)
$\lambda_i = 0.90$	0.87	1.52	57
$\lambda_i = 0.95$	0.88	1.55	60
$\lambda_i = 0.97$	0.88	1.56	60
$\lambda_i = 0.99$	0.89	1.58	61
fault +20%	0.88	0.95	19
fault +50%	0.88	1.56	60
fault +100%	0.88	2.50	106

5.2. Update frequency, measurement noise, and multistep-aware identification

The three studies of this subsection quantify the design tradeoffs analyzed in Sections 3.2 and 4.1. All of them use the nominal configuration of Section 5 with four seeds so that the first row of each table—$T_u = 1$, $\bar{n}_i = 0$, $\nu_i = 0$—is the nominal AK-DMPC and reproduces, at 1.30 cm, the eight-seed figure 1.35 ± 0.20 of Table 3. Aside from the closed-loop RMSE, we report the *eight-step network prediction error*, obtained by rolling the identified predictors of all three subsystems forward over the full horizon $H = 8$ along the realized inputs, each subsystem feeding its neighbors its own predicted lift; this is exactly the quantity $\delta(H)$ bounded by Theorem 4.3 with $\bar{\eta}_j = 0$.

Update frequency. Table 5 confirms Corollary 3.7 point by point. The closed-loop RMSE and the eight-step prediction error both degrade monotonically with T_u and saturate once the identifier is effectively frozen, and the fault recovery time is the quantity that suffers first and most, tripling already at $T_u = 5$. Against this, the computational saving is negligible: The identifier accounts for under 5% of the per-subsystem online cost and the quadratic program for the rest, so the entire saving available from intermittent updating is bounded by that 5%. On this benchmark, there is therefore no reason to update less often than every sample, and Corollary 3.7 explains why the conclusion is structural rather than accidental: The cost falls as $1/T_u$ from an already small base, and the residual radius grows affinely in T_u from a base that is the dominant error term.

Table 5. Effect of the identifier update period T_u (means over four seeds). The relative identifier cost is $1/T_u$ of the nominal, itself under 5% of the per-subsystem quadratic program cost.

T_u (samples)	Full-run RMSE (cm)	Post-fault RMSE (cm)	Recovery (samples)	8-step pred. error (cm)
1	1.30	1.56	60	2.07
2	1.62	2.02	114	2.62
5	1.84	2.32	218	3.25
10	1.89	2.36	210	3.56
20	1.92	2.38	182	3.80
50	1.93	2.37	182	3.96

Measurement noise. Table 6 adds zero-mean Gaussian noise of standard deviation σ_n to every level measurement used by both the identifier and the controller; the reported errors are still computed against the true levels. The degradation is graceful and close to affine in σ_n, as Corollary 3.8 predicts: A noise of 1 cm, roughly 4% of the operating levels, costs 17% of tracking accuracy, and even 2 cm—an implausibly poor level sensor—leaves AK-DMPC well inside the performance of the frozen Koopman baseline. The small-gain diagnostic is essentially unaffected, with $\rho(A_{\mathrm{cl}})$ moving from 0.51 to 0.49, so the certificate is not compromised by the noise at these levels. The dead zone behaves as analyzed: At $\varrho_i = 0.5\sigma_n$, it removes a quarter of the updates while very slightly improving the accuracy, so it is a free reduction of the estimator's duty cycle; at $\varrho_i = 2\sigma_n$, it removes more than 80% of them, and the adaptation it blocks is no longer only noise, so the tracking error rises above that of the unprotected estimator. The threshold must therefore be matched to the noise level and not chosen conservatively, which is the practical form of the tradeoff stated in item (ii) of Section 3.2.

Table 6. Measurement noise sensitivity (means over four seeds). σ_n is the standard deviation of the additive level noise seen by the identifier and by the controller; ϱ_i is the dead-zone radius on the measured output residual; "skipped" is the fraction of closed-loop updates suppressed by the dead zone.

σ_n (cm)	Dead zone ϱ_i	Full-run RMSE (cm)	Post-fault pred. error (cm)	$\rho(A_{\mathrm{cl}})$	Skipped
0	–	1.30	0.77	0.51	0%
0.25	–	1.32	0.80	0.51	0%
0.25	$0.5\sigma_n$	1.32	0.80	0.51	13%
0.25	$2\sigma_n$	1.32	0.81	0.50	46%
0.5	–	1.37	0.87	0.51	0%
0.5	$0.5\sigma_n$	1.37	0.87	0.50	19%
0.5	$2\sigma_n$	1.38	0.91	0.50	63%
1.0	–	1.52	1.07	0.50	0%
1.0	$0.5\sigma_n$	1.51	1.07	0.50	25%
1.0	$2\sigma_n$	1.55	1.15	0.50	75%
2.0	–	1.90	1.53	0.49	0%
2.0	$0.5\sigma_n$	1.90	1.52	0.49	28%
2.0	$2\sigma_n$	2.00	1.66	0.50	83%

Multistep-aware identification. Table 7 evaluates the amplification-penalized update (4.11). The mechanism works as Corollary 4.5 says it should: Increasing ν_i contracts the penalized blocks and lowers $\rho(A_{\mathrm{cl}})$ from 0.51 to 0.47 at $\nu_i = 0.005$, and the one-step fit improves slightly because the mild ridge also regularizes the estimate. What it does *not* do on this benchmark is improve the eight-step error, which stays at 2.07–2.08 cm before deteriorating for larger ν_i as the bias (4.12) takes over. Equation (4.9) explains the outcome quantitatively: With $\rho(M) \approx 0.5$, the amplification is only $\|\mathcal{A}_H\| \approx 2$, and the measured ratio between the eight-step and the one-step error is 1.9, so almost all of the multistep error is residual level, and almost none of it is amplification. There is consequently nothing for the

penalty to recover here, and the honest conclusion is that (4.11) is a tool for the regime it was derived for—networks whose identified predictors are weakly contractive or strongly coupled, where $\rho(M) \to 1$ and $\|\mathcal{A}_H\|$ grows without bound—and not a generic improvement. The value of (4.9) is precisely that it makes this diagnosis before the fact: The ratio $\|\mathcal{A}_H\|$ is computable online from the identified matrices, so a designer can tell whether the multistep error of their network is amplification-limited or residual-limited and reach for (4.11) only in the first case.

Table 7. Multistep-aware identification: Effect of the amplification penalty ν_i of (4.10)–(4.11) (means over four seeds). The amplification is the ratio of the eight-step to the one-step network prediction error.

ν_i	Full-run RMSE (cm)	1-step pred. error (cm)	8-step pred. error (cm)	Amplification	$\rho(A_{\mathrm{cl}})$
0	1.30	1.10	2.07	1.98	0.51
0.005	1.32	1.00	2.08	2.13	0.47
0.01	1.39	0.99	2.14	2.18	0.48
0.02	1.58	1.08	2.32	2.15	0.49
0.05	2.17	1.60	2.98	1.91	0.50

Remark 5.1. The decentralized linear controller is competitive only in the mild-coupling, near-equilibrium regime; once the nonlinearity and the interconnection are excited, it is the worst, consistent with the necessity of the cross-coupling terms $\hat{W}_{ij}$ in (2.3). The frozen Koopman controller matches AK-DMPC only while the plant coincides with the seeding data; the fault breaks that assumption, illustrating why online adaptation is required for drifting large-scale processes. Against the adaptive Hammerstein regulator, the message is subtler and, we believe, more honest: A well-tuned self-tuning regulator is a strong tracker, but it trades the guarantees and the smoothness for that accuracy, and it does not extend to constrained or certificate-bearing designs.

Remark 5.2. Three caveats delimit the results. (i) Persistency of excitation (Assumption 3) is enforced here by a small additive dither; for a purely regulating reference, it must be guaranteed by design, for example, through a vanishing exploration signal. (ii) The controller uses the full state x_i to form the lift $\Psi_i(x_i)$; output–feedback lifting from input–output data [61, 62] is left to future work. (iii) The forgetting factor λ_i trades adaptation speed against noise sensitivity through the term $\mu_i/(1-\lambda_i)$ of Theorem 3.3; $\lambda_i = 0.97$ was a robust compromise across all eight seeds, and Table 4 shows that the closed loop is in fact insensitive to this choice over $[0.90, 0.99]$ when the parameter variation is a jump rather than a continuous drift.

6. Conclusions

We have developed and certified an adaptive Koopman distributed predictive controller for large-scale interconnected nonlinear systems. Replacing the identified block-oriented model of classical self-tuning regulators by an online-estimated Koopman linear predictor with explicit cross-coupling and closing the loop with a distributed convex MPC, we proved that the adaptive identifier keeps a bounded covariance and drives the parameter error exponentially into a residual set fixed by the closure error and the drift, that the multistep prediction error admits an explicit bound, that the network is input-to-state stable

on an explicit region under a verifiable small-gain condition, and that recursive feasibility is robustly preserved. The asymptotic tracking error is governed by the dictionary richness and the parameter drift, and it vanishes as the dictionaries are enriched. Outside the slow-drift premise, a dwell-time corollary quantifies the bounded excursion and the geometric recovery produced by an abrupt parameter jump. The identifier was further analyzed under intermittent updating and bounded measurement noise and augmented with a convex, exactly recursive penalty on the amplification factor that converts a one-step residual into a multistep prediction error. A benchmark of three interconnected nonlinear tanks with an abrupt valve fault confirmed the analysis, with 28% and 58% RMSE reductions over frozen Koopman and decentralized linear controllers, parity in accuracy with an adaptive Hammerstein self-tuning regulator at one third of its control activity, and online verification of the bounded-closure and small-gain premises ($\rho(A_{\mathrm{cl}}) \approx 0.54$). Future work will address output–feedback lifting from input–output data [61]; dictionary learning with closure error certificates [27, 30]; a constructive synthesis of terminal ingredients valid over the whole projection set, which is the restrictive ingredient identified in Section 4.2; and resilience of the small-gain certificate to communication faults and adversarial inputs.

Appendix

A. Proof of Lemma 3.1

Write $Q_i(k) \triangleq P_i(k)^{-1}$ for the information matrix, and let

$$Q_i^{\mathrm{c}}(k) \triangleq \lambda_i Q_i(k-1) + \bar{\varphi}_i(k)\bar{\varphi}_i(k)^{\top} \tag{A.1}$$

denote the *candidate* produced by (3.1d) before the projection; by the matrix inversion lemma, (A.1) is the information form of the unprojected covariance update. Statement (ii) of the lemma is the assertion that $Q_i(k) = Q_i^{\mathrm{c}}(k)$ for $k > S_i$.

It is worth recording at the outset which route the proof does *not* take. Solving (A.1) in closed form gives

$$Q_i(k) = \lambda_i^k Q_i(0) + \sum_{t=1}^{k} \lambda_i^{k-t}\, \bar{\varphi}_i(t)\bar{\varphi}_i(t)^{\top}, \tag{A.2}$$

but (A.2) is valid only if the projection was inactive at *every* step between 0 and k, which is precisely what is in question: During the initial transient, the clipping may genuinely modify P_i, and a bound derived from (A.2) would then apply to a recursion that the algorithm does not run. The argument below therefore uses only two properties that survive an arbitrary amount of clipping.

(P1) The projection enforces the interval. By construction, $\Pi_{[\underline{p}_i, \bar{p}_i]}$ returns a symmetric matrix whose eigenvalues lie in $[\underline{p}_i, \bar{p}_i]$, so $\underline{p}_i I \preceq P_i(k) \preceq \bar{p}_i I$; that is, $\bar{p}_i^{-1} I \preceq Q_i(k) \preceq \underline{p}_i^{-1} I$ for every $k \geq 0$. For $k = 0$, this holds by the hypothesis $\underline{p}_i \leq p_{0,i} \leq \bar{p}_i$. This is statement (i).

(P2) Clipping is monotone. Raising the eigenvalues of P_i that fall below $\underline{p}_i$ and lowering those above $\bar{p}_i$ is, in the information coordinates, the operation of lowering the eigenvalues of Q_i above $\underline{p}_i^{-1}$ and raising those below $\bar{p}_i^{-1}$. In particular, whenever only the *upper* clip on P_i acts, the effect on Q_i is an increase in the positive-semidefinite order, $Q_i(k) \succeq Q_i^{\mathrm{c}}(k)$.

Step 1: The lower clip on P_i never acts. Suppose $Q_i(k-1) \preceq \underline{p}_i^{-1} I$, which holds for $k-1=0$ and, inductively, for every earlier step by (P1). Because $\|\bar{\varphi}_i\| < 1$, (A.1) gives

$$Q_i^{\mathrm{c}}(k) \;\preceq\; \lambda_i \underline{p}_i^{-1} I + I \;=\; (\lambda_i + \underline{p}_i)\underline{p}_i^{-1} I \;\preceq\; \underline{p}_i^{-1} I, \tag{A.3}$$

the last inequality being exactly the first design condition $\underline{p}_i \le 1-\lambda_i$ of (3.2). Hence, the candidate already satisfies $P_i^{\mathrm{c}}(k) \succeq \underline{p}_i I$, and the lower clip has nothing to do at any $k \ge 1$.

Step 2: The upper clip on P_i does not act for $k \ge S_i$. By Step 1, the only clipping that can occur is the upper one, so (P2) applies at every step and yields $Q_i(t) \succeq Q_i^{\mathrm{c}}(t)$ for all t. Chaining this inequality backwards through (A.1) over one excitation window and using $Q_i \succeq 0$,

$$Q_i^{\mathrm{c}}(k) \;\succeq\; \lambda_i^{S_i} Q_i(k-S_i) + \sum_{t=k-S_i+1}^{k} \lambda_i^{k-t}\, \bar{\varphi}_i(t)\bar{\varphi}_i(t)^{\top} \;\succeq\; \lambda_i^{S_i-1} \sum_{t=k-S_i+1}^{k} \bar{\varphi}_i(t)\bar{\varphi}_i(t)^{\top} \;\succeq\; \lambda_i^{S_i-1}\underline{\alpha}_i I, \tag{A.4}$$

where the middle inequality uses $\lambda_i^{k-t} \ge \lambda_i^{S_i-1}$ on the window, and the last one is Assumption 3. By the second design condition in (3.2), $\lambda_i^{S_i-1}\underline{\alpha}_i \ge \bar{p}_i^{-1}$, so $Q_i^{\mathrm{c}}(k) \succeq \bar{p}_i^{-1} I$; that is, $P_i^{\mathrm{c}}(k) \preceq \bar{p}_i I$: The candidate is already inside the interval, and the upper clip is inactive. This is statement (ii). Note that (A.4) was obtained without assuming anything about the clipping history before $k-S_i$; the transient is absorbed by (P2), which can only help the lower bound, and by (P1), which supplies the crude bound $Q_i(k-S_i) \succeq 0$ actually used.

Step 3. Steps 1 and 2 show that for $k > S_i$, neither clip acts, so $Q_i(k) = Q_i^{\mathrm{c}}(k)$, which is (3.3), and hence statement (iii). □

B. Proof of Lemma 4.6

(i) The quadratic stage cost gives the lower bound with constant $q_i = \lambda_{\min}(Q_i)$. The upper bound follows from the feasibility of the terminal controller κ_i^f, which provides an admissible suboptimal cost that is quadratic in $\left\|x_i - x_i^{\mathrm{r}}\right\|$, giving the constant $\bar{c}_i$ [58, Ch. 2]; the constant is uniform over the admissible data because these range over the compact projection set of Assumption 5.

(ii) By Assumption 4, $x_i^{\mathrm{r}} \in \operatorname{int} \mathbb{X}_i^f$, and $u_i^{\mathrm{r}} \in \operatorname{int} \mathbb{U}_i$. Both the terminal constraint and the box input constraints of $\mathbb{P}_i$ are therefore slack at $(x_i^{\mathrm{r}}, u_i^{\mathrm{r}})$, and as the optimizer of a strictly convex quadratic program depending continuously on the parameter, they remain slack on a ball $\mathcal{N}_i = \{s_i \le \eta_i\}$ of positive radius. On $\mathcal{N}_i$, the problem is an *unconstrained* strictly convex quadratic program in the decision $\mathbf{v}_i$; because the deviation cost (4.2) vanishes at the equilibrium $(x_i^{\mathrm{r}}, u_i^{\mathrm{r}})$, eliminating $\mathbf{v}_i$ leaves the single quadratic

$$V_i(x_i) = \tfrac{1}{2}(x_i - x_i^{\mathrm{r}})^{\top} \Pi_i (x_i - x_i^{\mathrm{r}}), \tag{B.1}$$

where $\Pi_i \succ 0$ is the constant condensed value Hessian, that is, the Schur complement of the reduced input Hessian $S_u^{\top} \bar{Q}_i S_u + \bar{R}_i$ in the full condensed Hessian of the stage cost. Hence, V_i is twice continuously differentiable on $\mathcal{N}_i$, with $V_i(x_i^{\mathrm{r}}) = 0$, $\nabla V_i(x_i^{\mathrm{r}}) = 0$ and $\nabla^2 V_i \equiv \Pi_i$ constant, so $\left\|\nabla^2 V_i\right\| \le L_i$ with $L_i = \|\Pi_i\|$, computed in closed form from the quadratic program data. Because no constraint is active on $\mathcal{N}_i$, no active-constraint multiplier enters ∇V_i, so this curvature bound is exact and renders the small-gain gains (C.10) verifiable in practice.

(iii) Condense (4.2) on $\mathcal{N}_i$, where it is unconstrained. Writing $\mathbf{r}$ for the stacked reference, the predicted output sequence is affine in the decision, and in the state, $\hat{\mathbf{x}}_i = S_u(\Theta)\mathbf{v}_i + f(x_i;\Theta,\boldsymbol{\zeta})$, so with $g \triangleq f(x_i;\Theta,\boldsymbol{\zeta}) - \mathbf{r}$, the optimal value is the explicit quadratic form

$$V_i(x_i;\Theta,\boldsymbol{\zeta}) = g^\top \Xi(\Theta)\, g, \qquad \Xi(\Theta) \triangleq \bar{Q} - \bar{Q}S_u(S_u^\top \bar{Q} S_u + \bar{R})^{-1} S_u^\top \bar{Q} \succ 0. \tag{B.2}$$

Both S_u and f are polynomial in the entries of Θ (they involve the powers $\hat{A}_i^\ell$, $\ell \le H$) and affine in $\boldsymbol{\zeta}$, and $\bar{R} \succ 0$ keeps the inverse in (B.2) bounded; hence, $\Theta \mapsto \Xi(\Theta)$ and $(\Theta,\boldsymbol{\zeta}) \mapsto g$ are continuously differentiable and therefore Lipschitz on the compact set of admissible data. Decompose $g = G(\Theta)(x_i - x_i^{\mathrm{r}}) + h(\Theta,\boldsymbol{\zeta})$, where h is the free-response offset evaluated at the reference; by Assumption 4, $h = 0$ when the data are exact, and the neighbors sit at their references, so $\|h\| \le c_h(\left\|\Theta - \Theta_i^\star\right\|_F + \sum_j \left\|\zeta_j - z_j^{\mathrm{r}}\right\|)$ for a constant c_h. Substituting this decomposition in (B.2) and subtracting the same expression for a second data pair, the difference collects three groups of terms: One proportional to s_i^2 times the data variation, one bilinear in s_i and the data variation, and one quadratic in the data variation. On $\mathcal{N}_i$ we have $s_i \le \eta_i$, so the first group is dominated by $\eta_i s_i$ times the data variation and merges with the second. Collecting the Lipschitz constants of G, Ξ, and h into $L_i^\Theta, L_i^\zeta, L_i^{(2)}$ yields (4.13). □

C. Proof of Theorem 4.8

Throughout, $s_i \triangleq \left\|x_i(k) - x_i^{\mathrm{r}}\right\|$ and $s_i^+ \triangleq \left\|x_i(k+1) - x_i^{\mathrm{r}}\right\|$, and $x(k) \in \Omega_\nu$, so $s_i \le \eta_i$ and Lemma 4.6(ii)–(iii) are available for every i. Three problems are compared at each instant: The *realized* problem $\mathbb{P}_i$ with data $(\hat{\Theta}_i(k), \boldsymbol{\zeta}_i(k))$, whose value is V_i^k; the *nominal* problem, with the exact parameter $\Theta_i^\star(k)$ and the neighbors held at their reference lifts z_j^{r}, whose value is V_i^{nom}; and the nominal problem propagated by its own predictor, whose successor state is $x_i^{\mathrm{nom}}(k+1)$.

Step 1 (nominal decrease). Let $\mathbf{u}_i^\star(k)$ be optimal for $\mathbb{P}_i$ and form the shifted candidate $\tilde{\mathbf{u}}_i = \{u_i^\star(1 \mid k), \ldots, u_i^\star(H-1 \mid k), \kappa_i^f(\hat{x}_i(H \mid k))\}$, feasible for the nominal predictor by Assumption 4. The terminal inequality of Assumption 4 telescopes the cost difference into

$$V_i^{\mathrm{nom}}(x_i^{\mathrm{nom}}(k+1)) - V_i^{\mathrm{nom}}(x_i(k)) \le -\left\|x_i(k) - x_i^{\mathrm{r}}\right\|_{Q_i}^2 \le -q_i s_i^2, \tag{C.1}$$

exactly as in the nominal MPC stability proof [57, 58]. Combining (C.1) with the sandwich of Lemma 4.6(i) also gives the bound $\left\|x_i^{\mathrm{nom}}(k+1) - x_i^{\mathrm{r}}\right\| \le \sqrt{\bar{c}_i/q_i}\; s_i$, used in Step 3.

Step 2 (decomposition). The quantity to be estimated is the change of the *time-varying* function V_i^k along the *realized* motion (Remark 4.1). Insert the nominal quantities and split it into four exactly compensating pieces,

$$\begin{aligned} V_i^{k+1}(x_i(k+1)) - V_i^k(x_i(k)) = {} & \underbrace{\left[V_i^{k+1}(x_i(k+1)) - V_i^{\mathrm{nom}}(x_i(k+1))\right]}_{\text{(a) data mismatch at } k+1} \\ & + \underbrace{\left[V_i^{\mathrm{nom}}(x_i(k+1)) - V_i^{\mathrm{nom}}(x_i^{\mathrm{nom}}(k+1))\right]}_{\text{(b) prediction error moves the argument}} \\ & + \underbrace{\left[V_i^{\mathrm{nom}}(x_i^{\mathrm{nom}}(k+1)) - V_i^{\mathrm{nom}}(x_i(k))\right]}_{\text{(c) nominal decrease } \le -q_i s_i^2} \end{aligned}$$

$$+ \underbrace{\left[V_i^{\mathrm{nom}}(x_i(k)) - V_i^k(x_i(k))\right]}_{\text{(d) data mismatch at } k}. \tag{C.2}$$

Term (c) is (C.1). Terms (a) and (d) are what Remark 4.1 announced: They exist only because V_i^k depends on $(\hat{\Theta}_i, \boldsymbol{\zeta}_i)$ and those data change between k and $k+1$. Term (b) is the one that (C.4) below quantifies.

Step 3 (term (b): The perturbation estimate). Let $e \triangleq x_i(k+1) - x_i^{\mathrm{nom}}(k+1) = C_i\delta_i(1 \mid k)$. Because V_i^{nom} is *exactly* the quadratic (B.1) on $\mathcal{N}_i$ (Lemma 4.6(ii)), the expansion

$$V_i^{\mathrm{nom}}(y+e) - V_i^{\mathrm{nom}}(y) = \nabla V_i^{\mathrm{nom}}(y)^\top e + \tfrac{1}{2}\, e^\top \Pi_i\, e \tag{C.3}$$

is an *identity*, not a truncated Taylor expansion: There are no higher-order terms to bound, and the "remainder" is the single, explicit quadratic $(1/2)e^\top\Pi_i e$. Taking $y = x_i^{\mathrm{nom}}(k+1)$ using $\left\|\nabla V_i^{\mathrm{nom}}(y)\right\| = \left\|\Pi_i(y - x_i^{\mathrm{r}})\right\| \le L_i\sqrt{\bar{c}_i/q_i}\, s_i$ from Step 1 and $\|\Pi_i\| \le L_i$,

$$\left|V_i^{\mathrm{nom}}(x_i(k+1)) - V_i^{\mathrm{nom}}(x_i^{\mathrm{nom}}(k+1))\right| \;\le\; L_i\sqrt{\tfrac{\bar{c}_i}{q_i}}\; s_i\,\|e\| + \tfrac{L_i}{2}\,\|e\|^2\,. \tag{C.4}$$

The validity of (C.4) is thus exactly the validity of Lemma 4.6(ii): It holds on $\mathcal{N}_i$, that is, for $s_i \le \eta_i$, and this is the origin of the regional nature of the theorem (Remark 4.9). It remains to bound $\|e\|$. Applying Theorem 4.3 at $\ell = 1$ with $\delta(0) = 0$ gives $\|\delta_i(1 \mid k)\| \le c_i(0)$, and hence,

$$\|e\| \;\le\; \|C_i\|\; c_i(0) \;\le\; \|C_i\|\left(\kappa_i d_i + \sum_{j\in\mathcal{N}_i} w_{ij}\bar{\eta}_j\right), \qquad \kappa_i \triangleq (1 + |\mathcal{N}_i|)\bar{z} + \bar{u} + 1, \tag{C.5}$$

with d_i as in (4.17). Note that (C.5) involves no neighbor deviation s_j: The coupling enters through terms (a) and (d), not through (b), which is what makes the two mechanisms separately visible.

Step 4 (terms (a) and (d): The data mismatch). Apply Lemma 4.6(iii) with $\Theta' = \hat{\Theta}_i$, $\Theta = \Theta_i^\star(k)$, $\boldsymbol{\zeta}' = \boldsymbol{\zeta}_i$, and $\boldsymbol{\zeta} = \{z_j^{\mathrm{r}}\}$. Set

$$D_i \triangleq \varepsilon_i + \varsigma_i, \qquad Z_i \triangleq \sum_{j\in\mathcal{N}_i}\left\|\zeta_j - z_j^{\mathrm{r}}\right\| \;\le \sum_{j\in\mathcal{N}_i}(L_{\psi,j}s_j + \bar{\eta}_j), \tag{C.6}$$

where the bound on Z_i uses the Lipschitz property of Ψ_j (Definition 2.1) and ς_i accounts, through Lemma 4.7, for the change of $\hat{\Theta}_i$ between k and $k+1$ so that a single D_i covers both (a) and (d). Then,

$$\text{(a)} + \text{(d)} \;\le\; (L_i^\Theta D_i + L_i^\zeta Z_i)(s_i + s_i^+) + 2L_i^{(2)}(D_i^2 + Z_i^2). \tag{C.7}$$

Step 5 (Young's inequalities and the gain matrix). Insert (C.1), (C.4)–(C.5), and (C.7) into (C.2). Every cross term is now of one of three types, and each is reconciled with the quadratic dissipation $-q_i s_i^2$ by Young's inequality:

$$\kappa_{ij}\, s_i s_j \le \varepsilon_{ij}\, s_i^2 + \frac{\kappa_{ij}^2}{4\varepsilon_{ij}}\, s_j^2, \qquad \kappa_{ij} \triangleq L_i^\zeta L_{\psi,j} + L_i\sqrt{\tfrac{\bar{c}_i}{q_i}}\,\|C_i\|\, w_{ij}L_{\psi,j}, \tag{C.8a}$$

$$\kappa_i^d\, s_i d_i \le \frac{q_i}{8}\, s_i^2 + \frac{2(\kappa_i^d)^2}{q_i}\, d_i^2, \qquad \kappa_i^d \triangleq L_i\sqrt{\tfrac{\bar{c}_i}{q_i}}\,\|C_i\|\,\kappa_i + L_i^\Theta, \tag{C.8b}$$

$$\theta_i\, s_i^+ \le \frac{\epsilon}{2}\,(s_i^+)^2 + \frac{1}{2\epsilon}\,\theta_i^2, \qquad \theta_i \triangleq L_i^{\Theta} D_i + L_i^{\zeta} Z_i. \tag{C.8c}$$

The last of these is the only step that needs comment. It produces a term in $(s_i^+)^2$, which refers to the successor state and therefore cannot be charged to the dissipation at time k, but by Lemma 4.6(i), $(s_i^+)^2 \le V_i^{k+1}(x_i(k+1))/q_i$, and this quantity appears on the left-hand side of (C.2). Choosing $\epsilon < 2q_i$ and moving $\epsilon V_i^{k+1}/(2q_i)$ to the left multiplies the whole inequality by $(1-\epsilon/(2q_i))^{-1}$, a factor that can be made arbitrarily close to one and is absorbed into the constants below. The quadratic remainders $L_i\,\|e\|^2/2$, $2L_i^{(2)} D_i^2$, and the $\bar\eta_j$ contributions are functions of the disturbance only and go into σ_i, whereas $2L_i^{(2)} Z_i^2 \le 4L_i^{(2)}|\mathcal{N}_i|\sum_j L_{\psi,j}^2 s_j^2 + 4L_i^{(2)}\bar\eta_\Sigma^2$ splits between the s_j^2 terms and σ_i. Setting $\sigma_i(d_i) \triangleq (2(\kappa_i^d)^2 q_i^{-1} + L_i\,\|C_i\|^2\,\kappa_i^2/2 + 2L_i^{(2)})d_i^2 \in \mathcal{K}_\infty$ and choosing $\sum_{j\in\mathcal{N}_i}\varepsilon_{ij} < 3q_i/4$ with margin $\alpha_i \triangleq q_i - q_i/4 - \sum_{j\in\mathcal{N}_i}\varepsilon_{ij} > 0$, the decomposition (C.2) becomes

$$V_i^{k+1}(x_i(k+1)) - V_i^k(x_i(k)) \le -\alpha_i s_i^2 + \sum_{j\in\mathcal{N}_i} g_{ij} s_j^2 + \sigma_i(d_i), \qquad g_{ij} \triangleq \frac{\kappa_{ij}^2}{4\varepsilon_{ij}} + 4L_i^{(2)}|\mathcal{N}_i| L_{\psi,j}^2. \tag{C.9}$$

By Lemma 4.6(i), $s_i^2 \ge V_i/\bar c_i$, and $s_j^2 \le V_j/q_j$; substituting into (C.9),

$$V_i^{k+1}(x_i(k+1)) - V_i^k(x_i(k)) \le -\frac{\alpha_i}{\bar c_i}\,V_i^k + \sum_{j\in\mathcal{N}_i}\gamma_{ij}\,V_j^k + \sigma_i(d_i), \qquad \gamma_{ij} \triangleq \frac{\bar c_i\, g_{ij}}{\alpha_i\, q_j}, \tag{C.10}$$

so each V_i^k is a local ISS Lyapunov function with the explicit interconnection gains γ_{ij} (and $\gamma_{ij} = 0$ for $j \notin \mathcal{N}_i$). These gains are computable from the identified coupling $\hat W_{ij}$, the weight Q_i, and the constants $L_i, L_i^{\Theta}, L_i^{\zeta}, L_i^{(2)}, L_{\psi,j}, \bar c_i$; the symmetric choice $\varepsilon_{ij} = q_i/(4|\mathcal{N}_i|)$ gives $\alpha_i = q_i/2$.

Step 6 (small gain and the region). Under (4.15), there exists $v > 0$ with $\Gamma v \prec v$ [47]. Consequently, $V^k(x) = \max_i V_i^k(x_i)/v_i$ is an ISS Lyapunov function for the whole network [46, 48, 51], and (C.10) gives $V^{k+1}(x(k+1)) \le \theta V^k(x(k)) + \max_i \sigma_i(d_i)/v_i$ for some $\theta \in (0,1)$ determined by $\Gamma v \prec v$ and $\{\alpha_i/\bar c_i\}$. Choosing $d^\star > 0$ such that $\max_i \sigma_i(d^\star)/v_i \le (1-\theta)c^\star$ makes the sublevel set Ω_v of (4.16) forward invariant, the estimate just derived is valid at every subsequent instant, and standard comparison arguments turn it into (4.18) with $\beta \in \mathcal{KL}$ and $\gamma \in \mathcal{K}_\infty$ assembled from θ, $\{\sigma_i\}$ and v. Finally, by Theorem 3.3, $\varepsilon_i \le \bar p_i(\bar r_i + \underline{p}_i^{-1}\mu_i)/(1-\lambda_i)$ asymptotically, and by Lemma 4.7, $\varsigma_i = O(\bar r_i + \mu_i)$, so $d_i = O(\bar r_i + \mu_i)$. Substituting into (4.18) and letting $k \to \infty$ ($\beta \to 0$) gives $\limsup_k \left\|x_i(k) - x_i^{\mathrm{r}}\right\| \le \gamma_i'(\max_j(\bar r_j + \mu_j))$, which tends to zero as the dictionaries are enriched ($\bar r_j \to 0$) under slow drift ($\mu_j \to 0$). □

D. Proof of Proposition 4.11

Because the only state constraint is the terminal one, robust recursive feasibility follows from a tube argument on the terminal ingredients alone [35, 57, 58]. Let $\mathbf{u}_i^\star(k)$ be optimal at k and form the shifted candidate $\tilde{\mathbf{u}}_i$ of Step 1 in Appendix C, appended with the prestabilizing terminal move. The nominal prestabilized predictor $\tilde{\mathbf{u}}_i$ respects the box input constraint and, by the robust terminal invariance of Assumption 5, steers the nominal terminal state into $\mathbb{X}_i^f$ despite any neighbor disturbance in $\mathbb{W}_i$. The realized trajectory deviates from the nominal one by the prediction error, which propagates through the

contractive $A_{K,i}$ ($\left\|A_{K,i}\right\| \leq a_i < 1$) and is bounded, in the worst case, by the geometric sum $\bar{\rho}_i$ of (4.19) (Theorem 4.3 with the neighbor mismatch $\bar{\eta}_j$ and the offset evaluated at the end of the horizon, $c_i(H-1)$, which is the largest of the horizon). Because the terminal set was tightened by exactly the ball $\mathcal{B}_i$ of radius $\bar{\rho}_i$, the realized terminal state lies in the untightened $\mathbb{X}_i^f$, and the appended terminal move remains in $\mathbb{U}_i$ for $\bar{d}^\star, \bar{\eta}_j$ small enough. Hence, $\tilde{\mathbf{u}}_i$ is feasible for $\mathbb{P}_i$ at $k+1$. Feasibility concerns only the existence of this candidate and is therefore unaffected by any discontinuity of the optimizer after the recursive-least-squares update. □

Author contributions

Tawfik Guesmi: Conceptualization, methodology, formal analysis, writing – original draft; Mourad Elloumi: Conceptualization, methodology, formal analysis, supervision, writing – review and editing; Khalid Alqunun: Software, validation, visualization, writing – review and editing; Omar Naifar: Investigation, formal analysis, validation, writing – review and editing. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

No generative AI tools were used in the preparation of this manuscript.

Acknowledgments

The authors thank the University of Ha'il, the University of Gafsa, the University of Kairouan, and Sfax University for the research facilities made available for this study.

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. K. J. Åström, B. Wittenmark, *Adaptive control*, 2 Eds., Dover Publications, Mineola, NY, 2008.
2. G. C. Goodwin, K. S. Sin, *Adaptive filtering prediction and control*, Prentice Hall, Englewood Cliffs, NJ, 1984.
3. P. A. Ioannou, J. Sun, *Robust adaptive control*, Prentice Hall, Upper Saddle River, NJ, 1996.
4. M. Elloumi, S. Kamoun, Parametric estimation of interconnected nonlinear systems described by input-output mathematical models, *Int. J. Autom. Comput.*, **13** (2016), 364–381. https://doi.org/10.1007/s11633-016-0956-8
5. M. Elloumi, S. Kamoun, Design of self-tuning regulator for large-scale interconnected Hammerstein systems, *J. Control Sci. Eng.*, **2016** (2016), 6769714. https://doi.org/10.1155/2016/6769714
6. M. Elloumi, S. Kamoun, Adaptive control scheme for large-scale interconnected systems described by Hammerstein models, *Asian J. Control*, **19** (2017), 1075–1088. https://doi.org/10.1002/asjc.1443

7. B. O. Koopman, Hamiltonian systems and transformation in Hilbert space, *P. Natl. A. Sci.*, **17** (1931), 315–318. https://doi.org/10.1073/pnas.17.5.315
8. I. Mezić, Spectral properties of dynamical systems, model reduction and decompositions, *Nonlinear Dynam.*, **41** (2005), 309–325. https://doi.org/10.1007/s11071-005-2824-x
9. M. O. Williams, I. G. Kevrekidis, C. W. Rowley, A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition, *J. Nonlinear Sci.*, **25** (2015), 1307–1346. https://doi.org/10.1007/s00332-015-9258-5
10. P. J. Schmid, Dynamic mode decomposition of numerical and experimental data, *J. Fluid Mech.*, **656** (2010), 5–28. https://doi.org/10.1017/S0022112010001217
11. J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, J. N. Kutz, On dynamic mode decomposition: Theory and applications, *J. Comput. Dynam.*, **1** (2014), 391–421. https://doi.org/10.3934/jcd.2014.1.391
12. M. Korda, I. Mezić, On convergence of extended dynamic mode decomposition to the Koopman operator, *J. Nonlinear Sci.*, **28** (2018), 687–710. https://doi.org/10.1007/s00332-017-9423-0
13. M. J. Colbrook, A. Townsend, Rigorous data-driven computation of spectral properties of Koopman operators for dynamical systems, *Commun. Pure Appl. Math.*, **77** (2024), 221–283. https://doi.org/10.1002/cpa.22125
14. S. L. Brunton, M. Budišić, E. Kaiser, J. N. Kutz, Modern Koopman theory for dynamical systems, *SIAM Rev.*, **64** (2022), 229–340. https://doi.org/10.1137/21M1401243
15. M. Budišić, R. Mohr, I. Mezić, Applied Koopmanism, *Chaos: An Interd. J. Nonlinear Sci.*, **22** (2012), 047510. https://doi.org/10.1063/1.4772195
16. A. Mauroy, I. Mezić, Y. Susuki (eds.), *The Koopman operator in systems and control: Concepts, methodologies, and applications*, Lecture Notes in Control and Information Sciences, Springer, International Publishing, Cham, 2020. https://doi.org/10.1007/978-3-030-35713-9
17. S. L. Brunton, B. W. Brunton, J. L. Proctor, J. N. Kutz, Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control, *PLoS One*, **11** (2016), e0150171. https://doi.org/10.1371/journal.pone.0150171
18. E. Kaiser, J. N. Kutz, S. L. Brunton, Data-driven discovery of Koopman eigenfunctions for control, *Mach. Learn.-Sci. Techn.*, **2** (2021), 035023. https://doi.org/10.1088/2632-2153/abf0f5
19. A. Mauroy, J. Goncalves, Koopman-based lifting techniques for nonlinear systems identification, *IEEE T. Automat. Contr.*, **65** (2020), 2550–2565. https://doi.org/10.1109/TAC.2019.2941433
20. M. Haseli, J. Cortés, Learning Koopman eigenfunctions and invariant subspaces from data: Symmetric subspace decomposition, *IEEE T. Automat. Contr.*, **67** (2022), 3442–3457. https://doi.org/10.1109/TAC.2021.3105318
21. B. Lusch, J. N. Kutz, S. L. Brunton, Deep learning for universal linear embeddings of nonlinear dynamics, *Nat. Commun.*, **9** (2018), 4950. https://doi.org/10.1038/s41467-018-07210-0
22. S. E. Otto, C. W. Rowley, Linearly recurrent autoencoder networks for learning dynamics, *SIAM J. Appl. Dynam. Syst.*, **18** (2019), 558–593. https://doi.org/10.1137/18M1177846
23. H. Shi, M. Q. H. Meng, Deep Koopman operator with control for nonlinear systems, *IEEE Robot. Autom. Lett.*, **7** (2022), 7700–7707. https://doi.org/10.1109/LRA.2022.3184036

24. M. Korda, I. Mezić, Linear predictors for nonlinear dynamical systems: Koopman operator meets model predictive control, *Automatica*, **93** (2018), 149–160. https://doi.org/10.1016/j.automatica.2018.03.046

25. S. Peitz, S. Klus, Koopman operator-based model reduction for switched-system control of PDEs, *Automatica*, **106** (2019), 184–191. https://doi.org/10.1016/j.automatica.2019.05.016

26. S. Peitz, S. E. Otto, C. W. Rowley, Data-driven model predictive control using interpolated Koopman generators, *SIAM J. Appl. Dynam. Syst.*, **19** (2020), 2162–2193. https://doi.org/10.1137/20M1325678

27. V. Cibulka, M. Korda, T. Haniš, Dictionary-free Koopman model predictive control with nonlinear input transformation, *SIAM J. Appl. Dynam. Syst.*, **24** (2025), 94–130. https://doi.org/10.1137/23M1597174

28. M. Wang, X. Lou, W. Wu, B. Cui, Koopman-based MPC with learned dynamics: Hierarchical neural network approach, *IEEE T. Neur. Net. Lear.*, **35** (2024), 3630–3639. https://doi.org/10.1109/TNNLS.2022.3194958

29. M. Ławryńczuk, Koopman operator-based multi-model for predictive control, *Nonlinear Dynam.*, **112** (2024), 9955–9982. https://doi.org/10.1007/s11071-024-09615-7

30. F. M. Philipp, M. Schaller, K. Worthmann, S. Peitz, F. Nüske, Error analysis of kernel EDMD for prediction and control in the Koopman framework, *J. Nonlinear Sci.*, **35** (2025), 92. https://doi.org/10.1007/s00332-025-10182-3

31. T. Chen, J. Shan, Koopman-operator-based attitude dynamics and control on SO(3), *J. Guid. Control Dynam.*, **43** (2020), 2112–2126. https://doi.org/10.2514/1.G005006

32. Y. Wang, T. Chen, S. He, Z. Wei, Z. Yin, J. Tayebi, Koopman-operator-based model predictive control of test mass capturing for drag-free satellite, *Space-Sci. Tech.*, 2026, 0568. https://doi.org/10.34133/space.0568

33. J. C. Willems, P. Rapisarda, I. Markovsky, B. L. M. De Moor, A note on persistency of excitation, *Syst. Control Lett.*, **54** (2005), 325–329. https://doi.org/10.1016/j.sysconle.2004.09.003

34. J. Coulson, J. Lygeros, F. Dörfler, *Data-enabled predictive control: In the shallows of the DeePC*, In: 2019 18th European Control Conference (ECC), 2019, 307–312. https://doi.org/10.23919/ECC.2019.8795639

35. J. Berberich, J. Köhler, M. A. Müller, F. Allgöwer, Data-driven model predictive control with stability and robustness guarantees, *IEEE T. Automat. Contr.*, **66** (2021), 1702–1717. https://doi.org/10.1109/TAC.2020.3000182

36. H. Zhang, C. W. Rowley, E. A. Deem, L. N. Cattafesta, Online dynamic mode decomposition for time-varying systems, *SIAM J. Appl. Dynam. Syst.*, **18** (2019), 1586–1609. https://doi.org/10.1137/18M1192329

37. O. Sayed, S. Lucia, *Recursive least squares-based identification for multi-step Koopman operators*, In: 2024 European Control Conference (ECC), 2024, 941–946. https://doi.org/10.23919/ECC64448.2024.10591201

38. Z. Zeng, H. Wang, Y. Shan, Data-driven Koopman operator-based model predictive control with adaptive dictionary learning for nonlinear industrial process optimization, *Mathematics*, **14** (2026), 1320. https://doi.org/10.3390/math14081320

39. A. Dittmer, B. Sharan, H. Werner, *Data-driven adaptive model predictive control for wind farms: A Koopman-based online learning approach*, In: 2022 IEEE 61st Conference on Decision and Control (CDC), 2022, 1999–2004. https://doi.org/10.1109/CDC51059.2022.9992829

40. K. Loya, P. Tallapragada, Online learning of Koopman operator using streaming data from different dynamical regimes, *IFAC-PapersOnLine*, **58** (2024), 90–95. https://doi.org/10.1016/j.ifacol.2024.12.016

41. T. de Jong, V. Breschi, M. Schoukens, M. Lazar, *Koopman data-driven predictive control with robust stability and recursive feasibility guarantees*, In: 2024 IEEE 63rd Conference on Decision and Control (CDC), 2024, 140–145. https://doi.org/10.1109/CDC56724.2024.10885975

42. P. D. Christofides, R. Scattolini, D. Muñoz de la Peña, J. Liu, Distributed model predictive control: A tutorial review and future research directions, *Comput. Chem. Eng.*, **51** (2013), 21–41. https://doi.org/10.1016/j.compchemeng.2012.05.011

43. R. Scattolini, Architectures for distributed and hierarchical model predictive control—a review, *J. Process Contr.*, **19** (2009), 723–731. https://doi.org/10.1016/j.jprocont.2009.02.003

44. B. T. Stewart, A. N. Venkat, J. B. Rawlings, S. J. Wright, G. Pannocchia, Cooperative distributed model predictive control, *Syst. Control Lett.*, **59** (2010), 460–469. https://doi.org/10.1016/j.sysconle.2010.06.005

45. L. Ljung, *System identification: Theory for the user*, 2 Eds., Prentice Hall, Upper Saddle River, NJ, 1999.

46. Z. P. Jiang, A. R. Teel, L. Praly, Small-gain theorem for ISS systems and applications, *Math. Control Signals Syst.*, **7** (1994), 95–120.

47. S. N. Dashkovskiy, B. S. Rüffer, F. R. Wirth, Small gain theorems for large scale systems and construction of ISS Lyapunov functions, *SIAM J. Control Optim.*, **48** (2010), 4089–4118. https://doi.org/10.1137/090746483

48. T. Liu, D. J. Hill, Z. P. Jiang, Lyapunov formulation of ISS cyclic-small-gain in continuous-time dynamical networks, *Automatica*, **47** (2011), 2088–2093. https://doi.org/10.1016/j.automatica.2011.06.018

49. E. D. Sontag, Smooth stabilization implies coprime factorization, *IEEE T. Automat. Contr.*, **34** (1989), 435–443. https://doi.org/10.1109/9.28018

50. Z. P. Jiang, Y. Wang, Input-to-state stability for discrete-time nonlinear systems, *Automatica*, **37** (2001), 857–869. https://doi.org/10.1016/S0005-1098(01)00028-0

51. A. Mironchenko, *Input-to-state stability: Theory and applications*, Communications and Control Engineering, Springer International Publishing, Cham, 2023. https://doi.org/10.1007/978-3-031-14674-9

52. K. H. Johansson, The quadruple-tank process: A multivariable laboratory process with an adjustable zero, *IEEE T. Contr. Syst. T.*, **8** (2000), 456–465. https://doi.org/10.1109/87.845876

53. D. S. Aponte, D. Tellez-Castro, *Koopman-based MPC for the quadruple tank system: An experimental comparison*, In: 2025 IEEE 7th Colombian Conference on Automatic Control (CCAC), 2025, 1–6. https://doi.org/10.1109/CCAC64704.2025.11259347

54. J. L. Proctor, S. L. Brunton, J. N. Kutz, Generalizing Koopman theory to allow for inputs and control, *SIAM J. Appl. Dynam. Syst.*, **17** (2018), 909–930. https://doi.org/10.1137/16M1062296

55. S. Klus, F. Nüske, P. Koltai, H. Wu, I. Kevrekidis, C. Schütte, F. Noé, Data-driven model reduction and transfer operator approximation, *J. Nonlinear Sci.*, **28** (2018), 985–1010. https://doi.org/10.1007/s00332-017-9437-7

56. S. Aranovskiy, R. Ushirobira, D. Efimov, J. Wang, Accelerated convergence with improved robustness for discrete-time parameter estimation, *Syst. Control Lett.*, **168** (2022), 105344. https://doi.org/10.1016/j.sysconle.2022.105344

57. D. Q. Mayne, J. B. Rawlings, C. V. Rao, P. O. M. Scokaert, Constrained model predictive control: Stability and optimality, *Automatica*, **36** (2000), 789–814. https://doi.org/10.1016/S0005-1098(99)00214-9

58. J. B. Rawlings, D. Q. Mayne, M. Diehl, *Model Predictive Control: Theory, Computation, and Design*, 2 Eds., Nob Hill Publishing, Santa Barbara, CA, 2017.

59. D. Q. Mayne, Model predictive control: Recent developments and future promise, *Automatica*, **50** (2014), 2967–2986. https://doi.org/10.1016/j.automatica.2014.10.128

60. A. Bemporad, M. Morari, V. Dua, E. N. Pistikopoulos, The explicit linear quadratic regulator for constrained systems, *Automatica*, **38** (2002), 3–20. https://doi.org/10.1016/S0005-1098(01)00174-1

61. P. Schmitz, T. Faulwasser, K. Worthmann, Willems' fundamental lemma for linear descriptor systems and its use for data-driven output-feedback MPC, *IEEE Contr. Syst. Lett.*, **6** (2022), 2443–2448. https://doi.org/10.1109/LCSYS.2022.3161054

62. T. Faulwasser, R. Ou, G. Pan, P. Schmitz, K. Worthmann, Behavioral theory for stochastic systems? A data-driven journey from Willems to Wiener and back again, *Annu. Rev. Control*, **55** (2023), 92–117. https://doi.org/10.1016/j.arcontrol.2023.03.005

AIMS Press